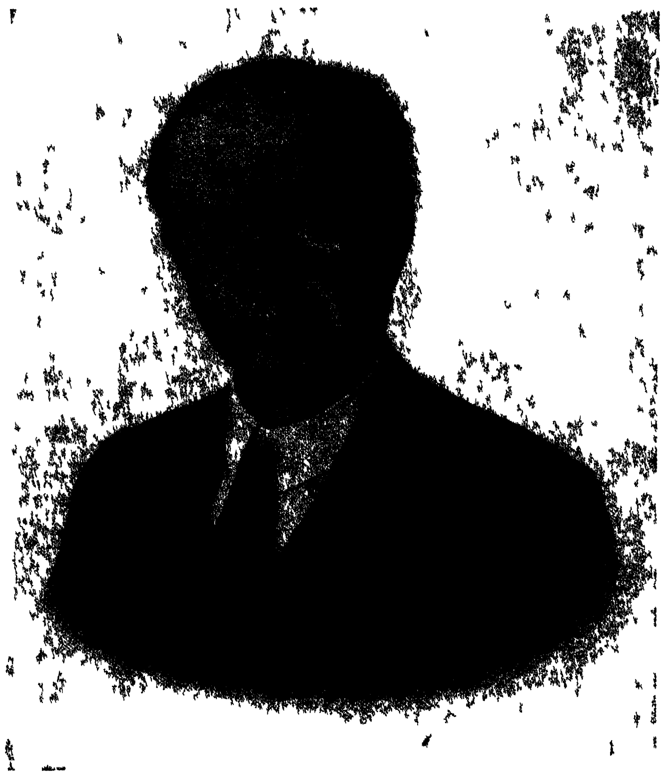
SHORT WAVE WIRELESS COMMUNICATION

INCLUDING ULTRA-SHORT WAVES



C. S. FRANKLIN.

Frontispiece.

SHORT WAVE WIRELESS COMMUNICATION

INCLUDING ULTRA-SHORT WAVES

By

A. W. LADNER

A.M.Inst.C.E.

Principal, The Marconi School
of Wireless Communication

and

C. R. STONER

B.Sc.(Eng.), A.M.I.E.E.

Late of Research Department, Marconi's
Wireless Telegraph Co., Ltd.
Lecturer at Queen Mary College

FOURTH EDITION

Revised and Enlarged

(Reprinted)

LONDON

CHAPMAN & HALL LTD.

37, ESSEX STREET, W.C.2

1946

First published, 1932
Second edition, 1934
Reprinted, 1934
Reprinted, 1935
Reprinted, 1935
Third edition, 1936
Reprinted, 1938
Reprinted, 1940
Fourth edition, 1942
Reprinted, 1943
Reprinted, 1944
Reprinted, 1946

REPRINTED FROM STANDING TYPE.
THE PAPER AND BINDING OF THIS
BOOK CONFORM TO THE AUTHORISED
ECONOMY STANDARDS

CATALOGUE No. 110/4

PRINTED IN GREAT BRITAIN BY JARROLD AND SONS, LIMITED, NORWICH

BOUND BY G. AND J. KITCAT, LIMITED, LONDON
Flexiback Binding.

Preface to First Edition

THE discovery, in the third decade of the present century, of the utility of "short" waves for world-wide communication, in which amateur workers can claim a large share of the credit, produced a revolutionary change in the science of wireless and had a profound influence, both technical and economic, on world communication.

The introduction of short wave beam methods enabled wireless, which had up to this time only carried a very small percentage of the world's long distance communications, to offer a high-speed telegraph service in most ways vastly superior to that of the older methods and at a very much smaller capital cost. In addition, it provided the only means so far developed for commercial trans-oceanic telephony.

In presenting a book on the principles of short wave wireless communication, it is the aim of the authors to fill an obvious gap in current literature and supply a text-book which shall satisfy the needs not only of engineers and telegraphists engaged in wireless, but cater for the scientific amateur and those who have already an outline knowledge of long wave working, such as may be gained by reading any elementary text-book on the subject.

Although the book deals especially with short wave communication, the field has not been restricted entirely to the peculiarities of short waves. A self-contained treatise has been aimed at, principles common both to long and short waves being introduced where the matter is necessary for the clearer explanation of the main subject.

The authors wish to pay tribute to the genius of C. S. Franklin, who from the earliest days of wireless has contributed so much to its progress in every field of action, and has played such a great part in the successful utilisation of short waves. His brilliant work in the development of the Beam and guiding it to success should alone be sufficient to ensure him an honoured

place in the history of communication engineering, and this represents but a small fraction of his contribution to the art.

In a book of this nature it is not possible to show adequate appreciation to our numerous friends who have supplied information, but we wish to make the following acknowledgments.

Firstly to Mr. Andrew Gray, until recently Technical General Manager of Marconi's Wireless Telegraph Co., Ltd., but for whose encouragement and kindly interest this book would not have been written. To Mr. T. L. Eckersley, who has been most helpful, and whose researches have supplied the bulk of the material for Chapter IV. To Mr. N. Wells, who has kindly checked the chapters on Transmitters, Feeders, Modulation and Aerials, and made many helpful suggestions. To Mr. Norman C. Stamford, who has assisted generally with the preparation of the book, and in the careful checking of the final manuscript.

To Marconi's Wireless Telegraph Co., Ltd., who have supplied much information and permitted the publication of the same, and to Mr. H. M. Dowsett for his light use of the Censor's pencil. Certain portions of the chapters on Modulation and Polar Diagrams of Aerial Arrays have previously appeared in the *Marconi Review*, and our thanks are due to the Editor for permission to republish. Thanks are also due for the loan of several blocks.

To the Imperial and International Communications Co., Ltd., for permission to publish information concerning their stations; particularly to Mr. J. Brown, Engineer-in-Charge of the Somerton Station, Mr. P. J. Woodward for information on the swinging of beams, and Mr. R. Keen for information regarding frequency measuring equipment.

To Standard Telephones and Cables, Ltd., for information supplied on their transmitters, array systems, and receivers.

To the Institute of Radio Engineers (U.S.A.) for permission to make use of papers which have appeared in their Journal, and to Messrs. Beverage and Carter, of the Radio Corporation of America, for checking the description of R.C.A. apparatus.

Our special thanks are due to Miss E. M. Elliot and Miss D. Turner for their great care in the preparation of the line drawings which illustrate the work. 1932.

Preface to Second Edition

IN presenting the second edition of "Short Wave Wireless Communication," we wish to thank those critics whose constructive comments have suggested to us where amendment and amplification of the work is desirable. We have found it unnecessary to alter either the chapter headings or their sequence, but new material has been included in Chapters II, III, IV, X, XI, XIII and XVII and an Appendix added in connection with feeder losses. In all some 36 pages of new material and 14 diagrams have been added.

Our thanks are due to Mr. T. L. Eckersley for kindly checking Chapters III and IV, to Mr. K. Tremellen for assistance in preparing a discussion on Shadow Charts, and to Mr. C. E. Rickard, O.B.E., for a considerable amount of original information on feeders and permission to publish data which he had collected.

Those chapters which deal with established theories and procedure contain a few selected references, but in cases where the chapter deals with controversial matter or very recent developments a representative selection of numbered references has been chosen.

January, 1934.

Preface to Third Edition

In the preparation of the third edition of "Short Wave Wireless Communication," much of the text has been revised and a new chapter on "Commercial Wireless Telephone Circuits" added. When preparing the previous editions the authors considered that practice in this field was scarcely stabilised sufficiently for incorporation in a textbook. This application of wireless has now become of such great importance and is of such great technical interest that a chapter upon it seemed very desirable.

In connection with this chapter the authors gratefully acknowledge the assistance they have received from the Engineer-in-Chief's Office of the British Post Office and especially from Mr. A. J. Gill. This acknowledgment extends also to the sections on the Frequency Checking Station and the Post Office Ultra Short Wave Circuits.

The chapter on aerials has been re-arranged and somewhat enlarged. In connection with this our thanks are due to Mr. T. L. Eckersley for his advice. Our indebtedness to Marconi's Wireless Telegraph Company acknowledged in the preface to the first edition, naturally increases as the succeeding editions are revised and enlarged.

Our thanks are due to Mr. T. D. Parkin for further information on quartz crystals and Mr. K. Tremellen for further information on ionosphere charts. To Messrs. C. and L. Kemp for details of commercial receivers, and to Mr. F. M. G. Murphy for information on Marconi Telephone Terminal Equipment and for helpful suggestions generally regarding the chapter on telephony. To the Institution of Electrical Engineers for permission to reprint Fig. 149 from their Journal. To the Bell System Telephone Labs. for permission to describe their Single Sideband System, and to Mr. H. Beverage for supplying and checking information on the R.C.A. apparatus. And to Miss Archer for the careful line drawings added to the new edition.

January, 1936.

Preface to Fourth Edition

In presenting the Fourth and war-time edition of Short Wave Wireless Communication the Authors regret that the exigencies of the times have prevented references to many new and interesting developments of short wave, and in particular ultra-short wave working. At the same time it has been found possible to add a great deal of new material, including some 180 new diagrams.

A change in the chapter sequence has been made, The increased use of line technique in many circuit problems suggested that line theory should be treated rather earlier in the book than formerly, and it appeared desirable that aerials should be more closely associated with the propagation of wireless waves.

The greater utilisation of ultra-short waves has necessitated that greater attention should be devoted to them, but since short and ultra-short waves have much in common, both are dealt with throughout in the text in appropriate chapters, instead as formerly of devoting one single chapter to ultra-short wave working.

The authors wish to record their indebtedness for information to the many members of the staff of the Marconi's Wireless Telegraph Company, whom it is not possible to mention individually. In particular, they wish to thank Mr. J. G. Robb, Technical Manager, and members of his staff for helpful comments, in particular Messrs. E. Green, S. B. Smith, W. H. Nottage and N. M. Rust. They also wish to acknowledge help received from Mr. G. Millington and Mr. J. Tremellen in checking Chapters IV and V; Mr. T. D. Parkin and Mr. D. Fairweather for general information on Quartz Crystals; Mr. E. C. S. Megaw for information on Electron Oscillators; Mr. R. B. Armstrong for information on Marconi Receiver Equipment; Mr. H. J. H. Wassell for information on Pentode Amplifiers; Mr. Wells for an interesting set of Polar Curves of

Horizontal Aerials; Dr. K. R. Sturley for help in preparation of the Reception Chapter; and Mr. R. L. Varney for details of Transmitter Equipment. Our special thanks are due to Mrs. E. Relf for the careful preparation of a large number of new line drawings. Thanks are due to the Institute of Radio Engineers for permission to publish curves shown in Figs. 107, 108 and 109, extracted from a paper by Brown on "Directional Antennas"; to the British Broadcasting Corporation for permission to publish the curves shown in Figs. 41 and 42; to the *Journal of Scientific Instruments* for permission to reprint the information contained in Table III, page 302; to Messrs. General Radiological, Limited, for supplying photographs of Dr. Schliepake's electrodes shown on page 528, and for information concerning their therapeutic apparatus; to Messrs. Stanley Cox, Limited, for information on their Inter-therm therapy set and to Messrs. Victor X-Ray Corporation for particulars of their Inductotherm therapy apparatus.

1942.

Contents

PAGE

APPENDICES

Selected References

Abbreviations Used:

B.S.T.J. . .	*Bell System Technical Journal.*
Elec. Com. .	*Electrical Communication.* (Issued by the International Standard Electric Corporation.)
Jour. A.I.E.E. .	*Journal of the American Institute of Electrical Engineers.*
J.Sc.I. . .	*Journal of Scientific Instruments.*
Journ. App. Phys.	*Journal of Applied Physics.*
J.I.E.E. . .	*Journal of Institution of Electrical Engineers.*
Mar. Rev. .	*Marconi Review.* (Issued by Marconi's Wireless Telegraph Co., Ltd.)
Phil. Mag. .	*The Philosophical Magazine and Journal of Science.*
P.I.R.E. . .	*Proceedings of Institute of Radio Engineers.* (Of America.)
P.O.E.E.J.	*Post Office Electrical Engineers' Journal.*
Proc. Roy. Soc. .	*Proceedings of Royal Society.*
Proc. Phys. Soc. .	*Proceedings of Physical Society.*
W.E. . .	*Wireless Engineer.* (*Formerly Experimental Wireless and Wireless Engineer.*)

Short Wave Wireless Communication

CHAPTER I

INTRODUCTION

SINCE the publication of the first edition of this book, the strangeness of working with very high frequencies has worn off and short and ultra-short waves are accepted as a usual and simple means of efficient communication. The passing years have consolidated our knowledge of the technique of transmitting and receiving apparatus, so much so, that modern short-wave gear is generally no more difficult to design and handle than its long-wave counterpart. Furthermore, the continued use of short waves has so increased our knowledge of the ionosphere that we now have a fairly comprehensive picture of its properties, though this knowledge has made us more conscious of its many vagaries.

It may be useful, at the beginning of this book, to review the requirements of any communication system. Intelligence cannot be transmitted by a single frequency, but needs an appropriate spectrum, the following table giving the frequencies involved in different types of signal.

TABLE I.

Type of Signal.	*Cycles/sec.*
High-speed Morse ...	0–200
Telephony	250–2,750
Broadcasting	30–10,000
Television	0–2,000,000

Note.—If an ordinary modulated transmission is used (see page 28) the frequency spectrum radiated for any type of intelligence will be double the maximum frequency involved.

It is interesting to observe that although the C.C.I.R. regulations suggest the staggering of adjacent broadcasting station carrier frequencies but 9 kc. apart, and receiver manufacturers design their receivers to admit 5 kc. only each side of the received wave, it is not possible for a manufacturing firm to sell a broadcast transmitter unless linear response is guaranteed to at least 8 kc. each side of the carrier.

Modern equipment has to be designed to pass the appropriate band of frequencies, preserving the relative amplitudes (and sometimes phases) within specified limits. Only a minor amount of energy must be produced at unwanted frequencies, and the output must be held constant at a correct level, in spite of changes in the intervening medium. These requirements are the same regardless of what carrier frequency is being used and are, indeed, the same for line working also.

A system may introduce distortion in three distinct ways:

(1) Frequency distortion (sometimes referred to as amplitude distortion), due to different frequency components of the signal being treated differently by the system, so that their relative amplitudes at the receiving end are not the same as initially produced at the transmitting end. This type of distortion must be kept small for all types of signal, except telegraphy.

(2) Phase distortion due to the relative phases of the various frequency components changing. This is immaterial when dealing with sound signals but is important for picture facsimile and television signals.

(3) Non-linear or harmonic distortion. If any part of the system has a non-linear characteristic, that is, the output is not strictly proportional to the input for all values of input met with, then new frequencies will be produced. These are harmonics of the various components of the signal and also frequencies which are the sum and difference between any two frequencies present in the signal. This type of distortion is serious at all times, and is particularly troublesome if one piece of equipment is handling two or more signals, as is frequently the case in line work and sometimes the case in wireless circuits. Any non-linearity in this case will result in interference between the two channels of communication.

An important feature in any channel is the ratio of signal

to noise. A strong signal is of little value if there are also unwanted disturbances present which produce an output comparable with the signal. A very weak signal having a high signal/noise ratio can, however, be made good use of if sufficient amplication be provided.

To show how short and ultra-short wave communication fits in, it may be helpful to consider a broad division of the wave range into the following arbitrary groups :

TABLE II.

Classification.	Frequencies. *Kilocycles.*	Wavelengths. *Metres.*
Long Waves	100–10	3,000–30,000
Medium Waves	750–100	400–3,000
Intermediate	2,000–750	150–400
Short Waves	30,000–2,000	10–150
Ultra-Short Waves	above 30,000	below 10

Of these only the long and short waves are useful for long distance services. The intermediate waves are limited in their range and hence are useful for serving a local area only ; for this reason they are used for broadcast purposes and are sometimes referred to as " broadcast waves."

Ultra-short waves have only a very limited reliable range which is very dependent upon the location of transmitter and receiver.

Since both long and short waves provide a possible means of long-distance communication, a summary of their relative merits is of interest.

In the case of long waves, although attenuation is not great, and the wireless path does not introduce distortion, the signal to noise ratio is poor (due to atmospherics), and everything possible must be done at the receiver to improve this ; but unfortunately many methods of so doing result in producing poor signal formation if carried too far. On the other hand, the attenuation remains approximately constant over long periods, and thus the signal arriving at the receiver is uniform in strength. Unfortunately the aerial is the weak link in the chain. It is easily possible to produce fairly efficiently the high-frequency currents required, but very difficult to get an

aerial circuit to radiate energy even with a costly mast and aerial system; since also the long-wave aerial cannot conveniently be made directive, even the small proportion of power radiated is not used to the best advantage for point-to-point working. The poor aerial efficiency reacts unfavourably on the transmitter, which must be of large power output, and hence it is difficult to control, besides being costly to build and operate. The modulation frequency is also limited by the time constants of the low-resistance aerials and closed oscillatory circuits, which are essential.

If short waves be used, however, the aerial system becomes much more efficient, and as the wavelength is small the whole aerial can be greatly reduced in size and cost, and also it now becomes practicable to design the aerial to concentrate the radiated energy into a beam, thereby making the best use of it. The transmitter can be greatly reduced in power, it is cheaper to construct and easier to operate, and may be keyed conveniently at high speeds, as at these very high frequencies the time constants of the circuits offer no bar to the highest modulation frequencies desired. The signal-noise ratio is very much better, so that services can be worked with field strengths at the receiving station of very low level, and the attenuation is perhaps rather less than for long waves, although this is governed by entirely different laws.

Unfortunately short waves are subject to rapid physical changes, so that the received signal varies over wide limits from moment to moment, and it is this fading phenomenon rather than atmospheric noise which is the controlling factor in economic design. For many classes of service, it is essential to provide the receiver with some form of compensation to correct for this and to deliver an approximately constant strength of signal irrespective of the fluctuations of input. For various reasons, which will be dealt with later, short waves do not form a distortionless connecting link, and in telephony and "picture telegraphy" the distortion may seriously interfere with reproduction.

Extremely marked diurnal and seasonal changes also occur, necessitating the use of different wavelengths at different times for the same service. In addition, channels working on routes passing near the magnetic poles are subject, during magnetic

storm periods, to such severe attenuation as sometimes to put them out of action.

Thus it will be seen that short waves offer the advantage of a cheap and efficient system, and a good signal-noise ratio, but suffer from a varying attenuation and a tendency for distortion.

Since the field strength used can be much lower than with long waves, the amplification needed at the receiver will be much greater and noises due to local electrical machinery and internal noises present new problems in receiver design.

It is of interest to observe that the chief contribution to early short-wave commercial communication was the "beam." The tremendous power gain of the first array systems used overcame the deficiencies of early transmitting and receiving apparatus and of the connecting link, the ether, and enabled telegraph circuits to operate at speeds not exceeded even to-day. Although the passing years have shown changes in "beam" design, first towards simplification with a view to economy of first cost, and more recently back to elaborate groups of smaller "beams," no advance in efficiency can be recorded. On the other hand, transmitting and receiving gear has improved in efficiency and design very considerably, very important developments being the perfecting of constant-frequency sources, and the technique of feeder and aerial circuit design. The power of short-wave transmitters has risen very considerably, and the employment of powers of 100 kW. and more has brought many problems to the designer.

The relative advantages of long and short wave working must now be discussed from quite another point of view. A consideration of Tables I and II will show that only eighteen telephone stations having a world-wide range could be accommodated within the whole band of long waves, whereas 5,090 such stations are possible on short waves. The whole long-wave band could not provide for one television transmission, if it was at all possible in other ways, and even the short-wave band would only accommodate seven. Actually, since the short waves are so useful, it would be out of the question to utilise 4 megacycles for one transmission (which would have a world-wide range, whether desired or not) and hence for such transmissions ultra-short waves are used. Besides the

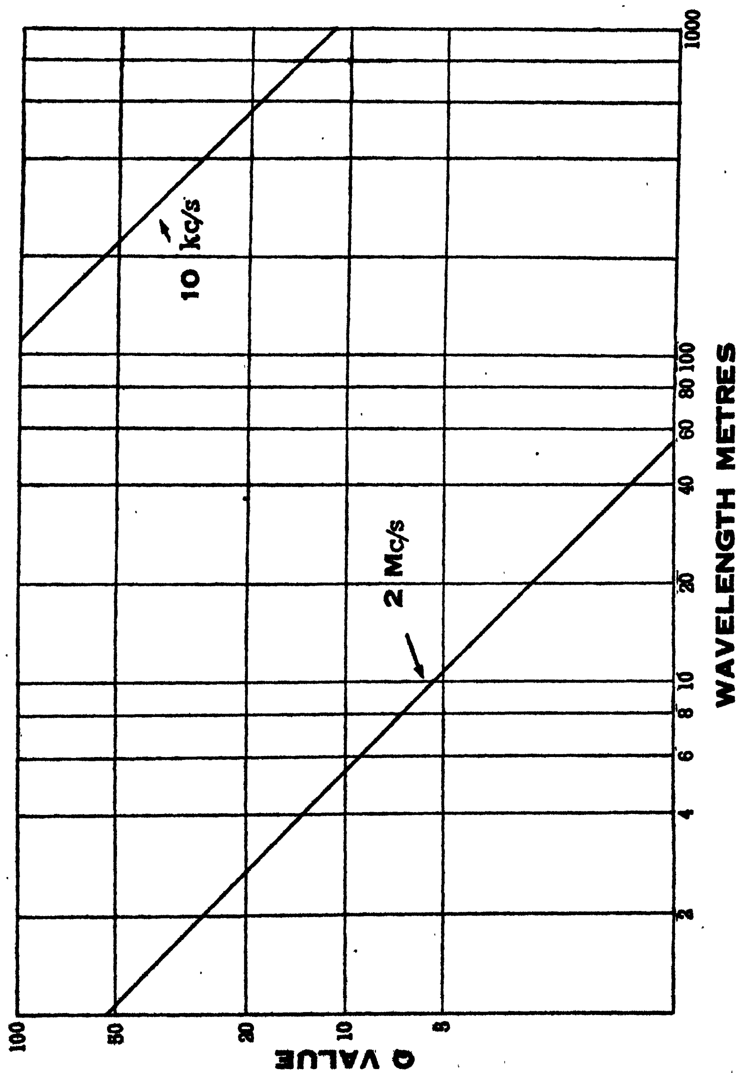

10 kc/s
2 Mc/s
WAVELENGTH METRES
2 4 6 8 10 20 40 60 80 100 1000
Q VALUE
100 50 20 10 5

FIGURE 1.

enormous frequency band available, these have the advantage that stations in different parts of the world may use the same frequencies without interference, since the range is limited.

Apart from the interference which would be created by transmitting television on anything but ultra-short waves, it would be found difficult to work on lower carrier frequencies owing to the modulation frequencies becoming a large percentage of the carrier frequency. This would mean that the tuned circuits would have to be less efficient, that is, their "Q" value ($Q = \omega L/R$ for a simple series circuit) must be smaller and this leads to difficulties in design.

For instance, assume that the highest modulation frequency to be dealt with in high-quality broadcasting is 10 kc./s. and in high-definition television 2,000 kc./s. Suppose that we can tolerate 2 db. attenuation of these highest frequencies. Then the Q value of any circuit in the radio-frequency tuning system at transmitter and receiver (including aerials) would need to be less than a value that will satisfy the above conditions, and limiting values for Q are shown in Fig. 1.

These curves show that as far as short waves and ultra-short waves are concerned, only the wide television spectrum is likely to be a determining factor in circuit design, as with other types of modulation circuit Q's of less than the limiting values shown would be adopted for other reasons.

CHAPTER II

A BRIEF HISTORY OF THE DEVELOPMENT OF SHORT WAVES

SHORT waves are now almost past history, but it is felt that a brief review of the development of short-wave wireless communication will still be of interest to the new student of the subject, in view of its revolutionary effect on wireless communication.

Commercial, long-distance, wireless communication may be said to have commenced in 1900 on a medium wavelength and, in the decade following, wireless progress was dependent upon medium and long waves. The longer the range desired the longer the wavelength used because the attenuation was thereby reduced.

After several years of experiment by a number of workers the principal laws governing communication on wavelengths from 200 metres upwards became known and wavelengths as long as 10,000 metres came into use for ranges of 5,000 miles and upwards. Empirical formulæ for attenuation were developed from measurements which were of necessity rather crude, but sufficiently good to show the manner in which the various factors affected the result. The attenuation was found to be less at night than by day, this effect being very marked on wavelengths of the order of 300 metres, and quite small on the longest wavelengths used. The attenuation, besides being less the longer the wavelength, was found to be dependent upon the nature of the earth's surface between the stations, being least for the best conductor, i.e., sea water, and owing to the bad results obtained with waves of 200 metres, shorter waves were not attempted for any except occasional very short range services. A wavelength of 120 metres (called tune A) was used, however, on ships from 1901 to 1909, and extraordinary night ranges were

found possible with very small power, 100 watts; with such waves, distances of 1,000 miles were frequently recorded, and 1,500 miles was recorded on more than one occasion, these ranges being obtained with a coherer receiver and tape recorder, a most insensitive device reckoned by modern standards.

In 1916 Marconi and C. S. Franklin recommenced experiments in Italy on very short waves, being attracted by the possibility of using directional transmission. It was thought that there might be uses for a short wave beam system, even if it had a very limited range. Special spark transmitters and parabolic reflectors were employed producing two or three metre waves, and a range of 6 miles over water was obtained, using a crystal receiver. The experiments were continued at Carnarvon in 1917 (on 4 to 5 metres), and a 20-mile range obtained when the transmitter was 600 feet above sea level, and the receiver visible from the transmitter. When both transmitter and receiver were placed at sea level the range fell to 4 miles, thus showing that the great attenuation in the neighbourhood of the transmitter was mainly produced by the earth. In 1919 valve transmitters were employed, the wavelength being subsequently raised to 15 metres, as this was considered to be the shortest at which the valves were likely to be satisfactory, and with an input of 200 watts, telephone signals were received at Kingstown, a distance of 70 miles.

In the following year a rotating parabolic 6-metre beam was fitted at Inchkeith, Firth of Forth, and run for 12 months, testing with the Royal Scot, a Leith to London coasting steamer. Experiments showed the utility of these waves for navigation purposes, and proved the abrupt falling off of signals outside the visual range, for this very short wavelength.

In 1921 a more ambitious short wave experiment was carried out. Valve transmitters and receivers with parabolic reflectors were set up near Hendon (N. London) and Birmingham, and a telephone circuit 97 miles long worked between them. Very strong signals on 15 metres were obtained with 700 watts input, the great utility of beam transmission being demonstrated. Much valuable experience in the design and operation of transmitters and receivers for these hitherto little used frequencies was gained. In particular, trouble occurred with the transmitting valves, due to the heavy oscillating currents

carried by the seals at these frequencies and also to local heating of the glass, and these troubles were traced and overcome.

Later in the same year a telephone circuit was set up between Southwold (Suffolk) and Zandvoort (Holland), using 100 metres without reflectors. The tests were very successful, and it was particularly noticed that on most nights and on some days good signals could be obtained at Oslo, Norway—a quite unexpected result at the time.

In 1922 the Marconi Co. continued short wave experiments definitely with a view to discovering their possibilities over long distances, and the historic Poldhu (Cornwall) Station was chosen for the purpose. Poldhu thus again became the site of epoch-making experiments, as it was from this station that the first transatlantic wireless signals were transmitted, 21 years previously.

A parabolic reflector system and half-wave aerial was erected for 97 metre transmission, this being considered to be the longest wavelength for which it was practicable to erect a reflector system.

In the spring of 1923 experimental transmission commenced from Poldhu, using an input power of 12 kW., receivers being installed on Marconi's yacht *Elettra*. The yacht sailed from Southampton, and it was found that signals up to St. Vincent (2,300 miles from Poldhu) were sufficiently good for commercial traffic purposes during nearly the whole 24 hours, although they were much stronger by night than by day. These results could be obtained with only 1 kW. input power at Poldhu. From the way in which the attenuation varied with distance it was deduced that a reliable commercial service could have been worked to Brazil during the hours of darkness, and the tests showed conclusively that waves below 100 metres were capable of giving reliable results over considerable distances, especially at night.

These results were of such importance that the British Government was officially informed of them by the Marconi Co., because the Government was embarking on the construction of very large and costly long wave stations for Imperial communications and it appeared likely that the short waves could be used with great advantage and economy.

Further trials took place from Poldhu in the spring of 1924, and on April 3rd telegraph signals from Poldhu were received in Sydney, N.S.W., and in May good telephone signals were also received. Successful and consistent reception was also reported from many other places throughout the world, and two important papers were read by Marconi before the Royal Society, in July and December, where the results were given and the flat projector for producing the beam described. In consequence of these results the Marconi Co. offered to enter into a very onerous contract with the British Government and the Dominions for the erection of a chain of short wave beam stations for Imperial communication. The terms were, briefly, as follows :

The Marconi Co. to erect in Great Britain short wave stations to communicate with Canada, Australia, India and South Africa. Communicating stations were to be erected in the above countries and were to be owned and operated by various commercial concerns under Government supervision. The British stations were to be bought and operated by the British Post Office, subject to their passing acceptance tests. These provided for seven days consecutive tests working duplex at not less than 100 words per minute, exclusive of repetitions to ensure accuracy, for 18 hours on the Canadian circuit, 11 hours on the South African, 12 hours on the Indian and 7 hours on the Australian circuits. The British Post Office could suffer no loss except some delay in proceeding with the much more costly long wave scheme, should the short wave beams fail, and the terms of the contract being acceptable to the Government, this was signed on July 28th, 1924.

The guaranteed traffic capacity was far greater than anything before known in wireless or ocean cable channels, and the general concensus of opinion at the time was that the " beam project " was but a speculative leap in the dark.

It is true that the evidence in 1924 demonstrated that all short wave signals were subject to violent fading and were, apparently, very erratic. These factors seemed to preclude their use for commercial high speed services, but Marconi believed that the beam would produce sufficient gain to overcome these defects. Such proved to be the case, and

Marconi deserves great credit for the correct interpretation of the scanty evidence. It was little realised at the time what a vital feature of a short wave circuit the directional array was to become.

All the available energies and all the accumulated experience of the Marconi Co. was now directed to the great enterprise. Another cruise was undertaken by the *Elettra* in August, 1924, especially to try the effects of still shorter wavelengths. It was quickly found that the daylight range increased greatly as the wavelength was decreased and daylight communication was achieved with Sydney on 32 metres. Meanwhile the construction of the beam stations was proceeding, research, design and construction having to proceed together.

The beam was obtained by the flat projector fed through concentric tube feeders developed by Franklin, and first tried out at Poldhu, and this took the place of the earlier parabolic reflector and single aerial, and was much more efficient. A very stable transmitter of 20 kW. input was developed and also a receiver, in the design of which the experience gained on the *Elettra* proved invaluable. Unexpected difficulties were met and conquered, and on October 25th, 1926, the Canadian circuit was opened, on a wavelength of 26·574 metres, the other circuits, Australia, South Africa and India, following at regular intervals during the next year, as shown in the chronological table at the end of the chapter.

It is interesting to note that the Australian circuit, Grimsby (England) to Rockbank (Australia), naturally the longest, was perhaps the least difficult to get going, and the phenomenal speed of 350 words per minute was worked during the acceptance tests. On this circuit the wavelength was not changed between the day and night conditions, as was necessary on the other circuits, but during certain hours, transmission took place over a great circle eastwards from Grimsby and at other hours over a great circle westwards. The opening of the beam stations placed British wireless circuits in the front rank of long distance communication systems and naturally attracted world-wide attention.

At this point it seems necessary to give a hint as to why the utility of short waves went so long undiscovered. All the early experiments had shown conclusively that in the neighbour-

hood of a very short wave transmitting station the attenuation was extremely great. What they had failed to reveal was that in addition to the surface wave which was being investigated and whose characteristics were understood, there was a radiation which entered the ionosphere and from there was bent towards the earth. If the bending was sufficiently great, the wave left the layer again and reached the earth's surface at some point generally very distant from the transmitter. Hence, if a portable receiver be carried away from a short wave transmitter, signals will at first be heard, but quickly become inaudible as the distance is increased. Then follows a zone in which only erratic signals can be received, but as the distance is still further increased strong signals will again be heard, and these signals, though varying rapidly in strength, may be well above noise level and quite suitable for commercial circuits. The intervention of the zone of unsatisfactory reception, or "skip zone" as it is termed, served to mask the existence of long distance waves. Thus in the London-Birmingham tests of 1921 on 15 metres, no doubt strong signals could have been received in many parts of the world, but no one thought of looking for them, and as no other reception on such wavelengths was being carried out no signals were received by accident.

The success of the Imperial beam stations was so complete that the cable services with which they were in competition lost a very great deal of traffic. It was realised that these cables were a great Imperial asset and it would be far better for the two methods of Empire communication to become complementary than that they should engage in competition which would bring serious financial embarrassment to the cable companies. In consequence, a Government-sponsored merger was planned and came into effect in 1928. As a result of this a company termed the Imperial and International Communications, Ltd., was formed to take over the Eastern Telegraph Co.'s network of cables, the Imperial Cables, the Imperial Beam Stations and all the Marconi Services. On the board of this company the British Government is represented, so that the interests of the British Empire are safeguarded and the public protected against exploitation by a monopoly.

We must now turn our attention to other organisations.

Naturally the Marconi beam project was watched with keen interest by radio engineers of other countries. The attitude of the American engineers towards short waves as a principal wireless channel was one of interested scepticism. They had little faith in their possibilities as carriers of high speed commercial traffic, although in September, 1924, the Radio Corporation of America decided to instal a 100-metre transmitter (WGH Tuckerton, N.J.) which might serve as an auxiliary to the long wave South American circuit, this being a difficult circuit on account of atmospherics. A simple "broadcast" aerial was used.

In the spring of 1925 other transmitters were brought into use in the same way. The wavelength used was gradually shortened, and in January, 1926, a transmitter was put into service on 14·9 metres for daylight service to South America. Operating speeds were still very low and the circuits in no way compared with those which came into use later. The American engineers at this time seemed doubtful as to the utility of beam aerials for short wave transmission and reception, and they did not try them. In consequence their signal-noise ratio was not nearly so good as on the Marconi circuits, and "echo" signals and excessive fading prevented high speeds being worked.

Whilst the R.C.A. had been investigating the behaviour of short waves by trying them out on actual commercial circuits, the American Navy, the Bell Telephone Laboratories and other organisations had been researching into their peculiarities. Also a short wave broadcasting station (KDKA) belonging to the Westinghouse Co. had been set up in 1925 at E. Pittsburg, transmitting on 62·7 metres, and it was successfully received and re-broadcast on several occasions in England.

The German Telefunken Co. had a long wave circuit to South America, which despite the use of very large power did not give very satisfactory results, and they were, therefore, soon alive to the importance of developing short wave services. They commenced with fairly long short waves and night working, but quickly reduced to as low as 15 metres. Their experiments were very successful and their stations (AGA and AGB), set up at Nauen, handled much traffic whilst still in the "makeshift" stage. These short wave stations began to

work traffic to South America early in 1925, and the traffic grew rapidly.

No directional arrangements were at first used and transmissions were at hand speed, but the behaviour of the circuits appeared to be reliable and consistent. Like the Americans, the Germans had no faith in the beam system, and it was not until after the British stations had fully proved the worth of the array aerial that Germany, America and France followed suit.

The activities of wireless amateurs in connection with short wave development must now be considered. Since the earliest days of wireless there have been a band of amateur workers of very varying capabilities. Some were amateur in the sense that wireless was to them only a hobby and not a profession, but their standard of knowledge was equal to that of many of the professional workers. Others were attracted by the novelty and interest of signalling to each other and cared little about scientific investigation. During the war the activities of amateurs of necessity completely ceased, though many of them found their knowledge useful in the Services, and when peace time conditions came again these amateurs naturally wished to recommence work and to use the greatly superior means now available. The number of stations engaged in serious work had, however, so greatly increased that it was realised amateur activities would have to be seriously curtailed and carefully controlled. In the British Isles it was finally decided that waves below 200 metres might be used, as these waves were not occupied by commercial or Service stations and were, indeed, considered useless except for very short ranges. The amateurs protested considerably against this decision, but were compelled to accept it.

It had long been known that occasional very long ranges were obtained with low power on wavelengths below 300 metres, in spite of the general ineffectiveness of these waves, and most operators were in the habit of repeating stories about freak results on such waves Spurred on by a few such results, with commendable pluck the amateurs of Great Britain and America organised a 200-metre transatlantic test in December, 1921. As the power restrictions of the American amateurs were more liberal than those of the British, the experiments were made only in the eastward direction, and it

is to be recorded that successful communication was obtained on several occasions.

In November, 1923, a French amateur established two-way communication by night on 100 metres with the U.S.A., and in January, 1924, British and U.S.A. amateurs were in communication under similar conditions. Before the end of this year the longest distances on the earth's surface had been covered by amateur transmissions, for on October 18th Mr. Goyder, of Mill Hill School, London, exchanged messages with Mr. Bell, of Dunedin, New Zealand, and others soon succeeded in maintaining contact both with New Zealand and Australia. These remarkable results, obtained by persons using very small power, and to whom wireless was merely a hobby, deserve the greatest praise for the ingenuity, resource and perseverance shown. Amateurs were led to try shorter waves and daylight communication, and as the technique of the very high frequencies was mastered, success came in February, 1925, when daylight transatlantic communication was maintained daily for a month.

On several occasions when the official and commercial short wave stations were still very few and the requirements for obtaining consistent results not understood, the amateur stations performed a very valuable service by maintaining contact with expeditions, etc., which were out of touch by other means of communication with civilisation. Thus in 1925 the Mill Hill station kept contact with an Arctic expedition when other means failed, and in the same year another English amateur (2NM) performed the same service for the Hamilton-Rice expedition in the wilds of Brazil.

The next great development of which we shall speak is the setting up of long distance telephone circuits by means of short waves. We have seen that experiments in short wave telephony took place very soon after telegraph signals had been sent over long distances, and these tests were quite successful. Much research had to be done, however, before commercial circuits could be brought into operation. What was desired was that the telephone networks in the various continents should be linked together by circuits which should be as reliable and give as good a standard of speech as long trunk lines. The immense difficulties in the way of adapting even the modern, greatly

improved submarine cable for transatlantic telephony and the colossal expense that would be involved pointed the way to the use of wireless.

The problem of wireless telephony first engaged the attention of the British Post Office in 1923 and in co-operation with the American Telephone and Telegraph Corp., succeeding years were spent in investigating the possibilities of transatlantic telephony on long wavelengths. These investigations resulted in a successful commercial channel being opened in 1927 on 6,000 metres.

When short waves came into use for telegraph services there was very much to be done before telephone circuits could be brought into operation. The chief troubles were due to "fading," and evidently if the short wave link is to be connected through ordinary land line circuits to ordinary subscribers the varying attenuation must be compensated for in some way, and various devices in the receiver now enable this to be done with some measure of success.

Because of the pioneer work done by the Marconi Company on the short wave beam services, the newly formed Imperial and International Communications Company anticipated that they would be able to participate in the extension of telephone communication for which the short wave Marconi beam was so suitable. In fact shortly after the inception of the telegraph service the Marconi Company developed a multiplex system by which the same aerial, transmitter and receiver could be used simultaneously for two telegraph channels and a telephone channel; but the British Government decided that wireless telephony in England should remain the monopoly of the British Post Office and the latter preferred to develop their own short wave telephone services independently of the Communication Company just formed.

Short wave channels developed in conjunction with the International Telegraph and Telephone Corp. were first brought into use to supplement the New York service, and during 1930 circuits were opened to various liners on the North Atlantic route, to South America and to Australia. Further services were opened at intervals until by 1934 there were twelve such channels operating from the Radio Section of the International Exchange in London.

At the English end the transmitters are all concentrated at Rugby and most of the receivers at Baldock (Hertfordshire). Aerial arrays of many different types are used at both transmitting and receiving stations.

As regards developments abroad, besides the distant ends of the circuits to London already mentioned, there are circuits in operation between New York and South America, and also to liners at sea. Buenos Aires has also circuits to Madrid, Paris and Berlin as well as London.

By these remarkable developments of the last year or so, the telephone subscriber in Britain is potentially connected to a large proportion of the subscribers in the world, and can carry on a conversation with a subscriber in Australia with almost as much reliability as with one in a town a few miles away. Further, he may get into communication with the passengers of most of the largest transatlantic liners.

In addition to long distance telephone channels, short waves have been utilised to provide a broadcasting service over long distances. The early American work has already been briefly noted. In 1927 an experimental broadcasting service was commenced from Chelmsford, England, and this continued until a permanent Empire Broadcasting Station was opened at Daventry at the close of 1932, by means of which all parts of the British Empire receive broadcasting from England at some suitable time during each twenty-four hours.

After the short waves had been well established, attention was again directed to the ultra-short waves and in the last few years they have in their turn been developed into a reliable means of communication, suitable, of course, only for short distances. A regular traffic circuit employing the shortest wavelength (18 cm.) is that first set up in 1931 by Standard Telephones and Cables, Ltd., across the Straits of Dover. The British Post Office installed, at the close of 1932, a circuit 12 miles long and working on 5 metres to provide an alternative telephone channel to the roundabout line circuits through the Severn Tunnel. More recently, the G.P.O. have made extensive use of ultra-short waves for linking outlying islands with the mainland of Great Britain, and also for providing additional circuits to Northern Ireland. In Italy Marconi and Mathieu produced perfectly stable duplex telephone

circuits and obtained ranges considerably in excess of the optical range, and a telephone circuit between the Vatican and the summer residence of the Pope was inaugurated in February, 1933, working on a wavelength of 60 cms. The Radio Corporation of America has installed a complete network of ultra-short wave stations linking together the various Hawaian islands.

Most broadcasting organisations have at various times carried out experimental transmissions on wavelengths of about 5 metres to supply a limited area with ultra high quality broadcasting without interference being caused over a much larger area.

More recently in the U.S.A. broadcast transmission on ultra-short waves employing wide deviation, frequency-modulation has been commenced, but so far the B.B.C. have not considered such a system will help much with the particular problems of broadcasting in this country. The transmission of television which, because of its high modulation frequencies, must also be carried out on ultra-short waves, has passed the experimental stage, and in 1936 the first regular television service was inaugurated by the B.B.C. from the Alexandra Palace, using a wavelength of 6.67 metres (45 Mc/s) for the vision transmission and 7.23 (41.5 Mc/s) for the sound transmission, the service having an effective entertainment range of some 50 miles radius.

Television transmissions are also made from New York, Paris, and Berlin.

Mention should also be made of the use of the higher wireless frequencies for medical purposes. Although frequencies of the order of one megacycle have been used for a number of years for diathermy work and to some extent in surgery, the use of higher frequencies enables a greater facility of application, a more uniform treatment and the possibility of reaching deep-seated tissue.

Thus we have sketched in outline the amazing and completely unexpected developments of short wave working. Perhaps in no department of applied science has there been such a surprising "kink" in the ordinary lines of progress. For so many years the radio engineer considered he understood at least in outline the behaviour of electro-magnetic waves,

only to find that there was an unknown set of phenomena of extreme usefulness at, so to speak, his very door.

In tracing out the rather involved story it will appear to have been written rather from the British point of view, and it may be thought that the work done in other countries has been dismissed too briefly, but it is a fact that the credit of early short wave development, and in particular beam working, must be given to Great Britain. It will be realised also that in a sketch of this length it is not possible to include all the outstanding items in the tale of progress.

CHRONOLOGICAL TABLE SHOWING THE DEVELOPMENT OF SHORT WAVES

1887 Original Hertz experiments.

1894 Lodge's demonstration at British Association meeting, Oxford.

1896 Beam demonstrated over 2 miles by Marconi.

1916 Marconi resumes short wave experiments.

1920 Inchkeith rotating beam set up.

1921 Aug. Southwold-Zandvoort 100-metre telephone tests.
London-Birmingham tests.
200-metre amateur transatlantic tests.

1922 Franklin's paper to I.E.E.

1923 Spring and summer. Transmissions between Poldhu and Marconi's yacht. Persistence of beam at great distances proved.

1923 Nov. Two-way amateur communication by night between France and U.S.A. on 100 metres.

1924 Jan. First two-way amateur communications between England and U.S.A. on 100 metres.
Spring. Further tests between Poldhu and Márconi's yacht.
April 3rd. Telegraphy to Sydney, Buenos Aires.
May. Telephony, Poldhu to Sydney.
July 2nd. Marconi's lecture to Royal Society.
July 28th. Beam contract signed.
Aug. Renewal of Poldhu tests in order to secure good daylight communication.

1924 Sept. Radio Corporation of America instal their first S.W. transmitter.

Oct. Complete tests from Poldhu to most parts of the world in daylight, using 32 metres.

Oct. 18th. First two-way communication between England and New Zealand on about 100 metres established by Messrs. Goyder and Bell.

1925 April. German short wave stations start working to South America.

Feb. First amateur 18–23 metre transmissions.

Aug. Experimental transmissions from Chelmsford commenced.

Contact by English amateur (2NM) with Hamilton-Rice expedition in Brazil, otherwise cut off from civilisation.

French Naval vessel, *Jaques Cartier*, conducts short wave tests.

1926 July. Marconi short wave transmitter at Ongar commenced working.

Oct. 25th. Canadian beam circuit opened.

1927 April 8th. Australian beam circuit opened.

July 5th. South African beam circuit opened.

Sept. 6th. Indian beam circuit opened.

Nov. 5th. 5SW Chelmsford short wave broadcasting station opened.

From this time onward developments in every country were great and no attempt will be made here to chronicle them. The following items are merely selected as being of special interest:

1929 Oct. Madrid-Buenos Aires telephone circuit (set up by the International Telephone and Telegraph Corporation) was opened.

1930 British Post Office telephone circuits to Sydney, Buenos Aires and ships at sea opened. Telephone circuits from Buenos Aires to Paris, Berlin and New York opened.

1931 March. Demonstration of 18 cm. waves by Standard Telephones.

1932 Hawaian telephone circuits opened.
Feb. Telephone service to South Africa opened.
Feb. Five metre telephone circuit across Bristol Channel erected.
Dec. Empire Broadcasting Station opened.
1933 Feb. Vatican 60 cm. circuit inaugurated.
May. Telephone service to India opened.
1936 Television commenced.

CHAPTER III

THE MODULATION OF HIGH FREQUENCY WAVES

COMMUNICATION systems in general are associated with the transmission and reception of waves whose complexity of form and frequency depend upon the type of intelligence being communicated. The effective frequency band of this intelligence varies from a few cycles per second, in the case of low-speed, Morse signalling to megacycles per second, as required for television.

One may conceive any communication system as having two essential features, each complementary to the other.

(1) A means of transferring energy from transmitter to receiver; thus we have sound waves in the case of speech between individuals; line currents in telegraphy and telephony; and ether waves in the case of wireless.

(2) A means of modulating this energy to conform to the desired intelligence, called the "signal." *

Systems may for convenience be divided into three classes:

(1) Non-carrier systems.
(2) Carrier systems.
(3) Suppressed carrier systems.

A non-carrier system is one in which energy transference only appears with the signal. If, however, transmitter and receiver are linked, even when no signal is being made, we have a carrier system. We have yet another case where, although no carrier links transmitter and receiver, it originally existed but has been suppressed in the channel, to be incorporated at the receiver, and such would be called a suppressed carrier system.

* Throughout this chapter we shall refer to the modulating component as the "signal."

Let us first discuss land line telegraphy, considering it from perhaps a rather unusual point of view. A Morse telegraph signal consists of three elements, dot, dash and space, but in the only systems we shall consider, the dot and dash are both formed by pulses of current in the same direction but of different duration, and will both be classed under the term "mark." One element of a signal will, therefore, consist of a mark and a space.

There are two possible ways of transmitting such an element illustrated in Fig. 2 (a and b). The signal can be made by a transmitter sending a current of amplitude "a" say, along

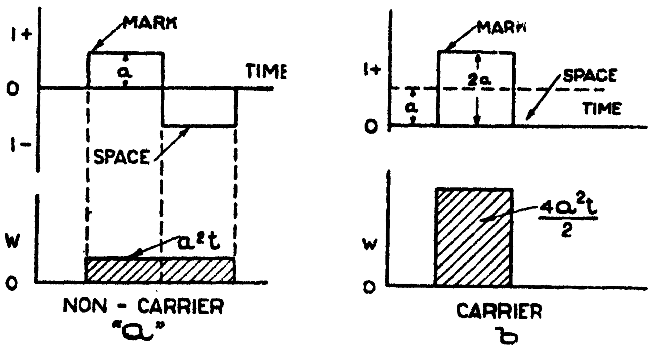


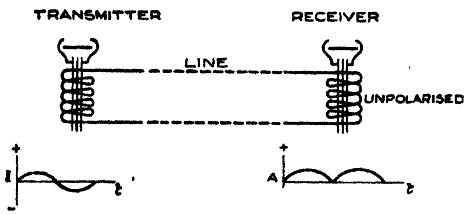
FIGURE 2.

a line, first in one direction and then in the other, giving an effective difference between mark and space of 2a. In this case we can say that the datum line of signal coincides with zero line of current in the system, and it is to be observed that a reversal of phase takes place at the half cycle of signal. This is an example of a non-carrier system, and this changing line current, which forms the signal, can be detected by a suitably designed receiver. This system is the usual "double current" telegraphy.

Carrier Systems. The second method is shown in Fig. 2b, and we may consider that a carrier current at least equal to "a" flows along the line and is augmented and reduced (but not reversed) by the signal. If the carrier is equal to the signal amplitude, the line current on mark is 2a and on space is zero. In this case the current always flows along the line

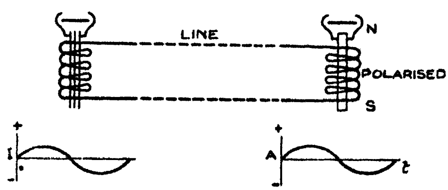
in the same direction and the carrier now forms a datum line for the signal wave.

Considering the power expended, it will be seen that in the first case it is half that of the second for the same amplitude signal, in spite of the fact that in the former current flows for both mark and space, and only during mark in the latter. It would appear from this that a signal can be transmitted along a communication channel with the least expenditure of energy when the datum line of signal coincides with zero line of current. Also that the signal made and received depends only upon the modulation component, and that the carrier contributes nothing to the intelligence except to form a datum line for the signal wave. The use of the carrier clearly involves an additional expenditure of energy which may be considerable.



NON CARRIER SYSTEM
UNPOLARISED RECEIVER

FIGURE 3.



NON CARRIER SYSTEM
POLARISED RECEIVER

FIGURE 4.

A carrier current, or its equivalent is, however, often necessary because it aids detection at the receiver.

Compare the two line cases, just mentioned; in the non-carrier example, for the receiver to detect the signal it has to be capable of following not only change of amplitude of current, but change of direction as well. On the other hand, if a carrier system is being used, the receiver has only to be able to detect change of amplitude and not change of direction.

For instance, consider a line telephone; if an unpolarised receiver is used, it can only interpret change of amplitude of current and not change of direction, for any rise of current, no matter what its direction, attracts the diaphragm, and the received signal appears distorted as shown in Fig. 3. If,

however, the receiver is polarised with permanent flux, giving an initial bias to the diaphragm, current in one direction augments the flux, and in the other direction decreases it, and in consequence a correct reproduction results as shown in Fig. 4. Had one of necessity to use the original type of detector it could be made to interpret both change of amplitude and direction of current by adding a D.C. carrier as shown in Fig. 5, although this carrier need not pass along the communication channel, but could be localised at the receiver circuit as shown in Fig. 6.

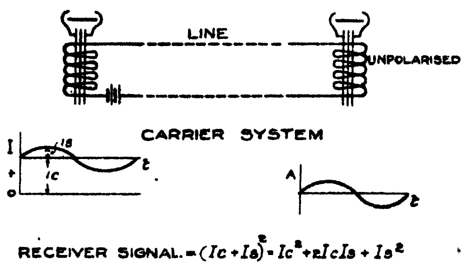


FIGURE 5.

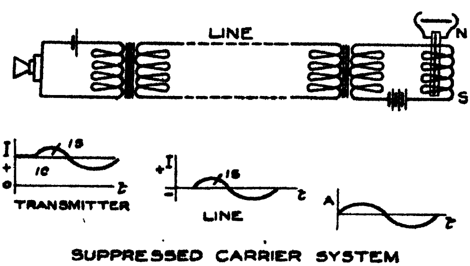


FIGURE 6.

Wireless communication is essentially a carrier system, because high frequencies are necessary for the effective transmission of the electro-magnetic wave. Thus instead of a D.C. we have an A.C. carrier creating an image wave (see examples in Figs. 8, 9 and 10), and because of this image wave, rectification is essential at the receiver, for the purpose of extracting the modulation.

It is of interest to consider the relationship of this A.C. carrier to the modulating component, and to discuss the possibilities of carrier suppression.

It may be pointed out here that the modulation of a carrier, whether D.C. or A.C., by a signal is often expressed by a modulation factor k, or if multiplied by 100 as a modulation percentage. The modulation factor expresses the ratio of signal amplitude to carrier amplitude and thus becomes unity if the carrier is fully modulated. For purposes of comparison, modulation measurements must always be made with a sinusoidal signal.

Modulation of a High-frequency Carrier. A sinusoidal carrier current which is of constant amplitude, frequency, and phase, may be expressed thus :

$$i=A \sin (\omega_c t+\phi) \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Where A =amplitude of the carrier current.
f_c=frequency of the carrier current.
$\omega_c=2\pi f_c$.
ϕ=phase angle of wave, relative to some arbitary datum.

If we bring about a periodic change of either amplitude, frequency, or phase at a rate which is slow compared with the carrier frequency, the process is called modulation. The change must be made about a datum having the same value as the quantity (amplitude, frequency, or phase) had in the unmodulated carrier.

Amplitude Modulation. For instance, suppose we wish to modulate the amplitude of the carrier by a low-frequency sine wave $A_s \sin \omega_s t$. If $A_s=k_a A$ (where $k_a \times 100$ is the percentage modulation), then A in expression (1) becomes :

$$A\ (1+k_a \sin \omega_s t) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

and it will be seen that this indicates a periodic amplitude change about the value (A) of the unmodulated amplitude. If we had merely changed A into $k_a \sin \omega_s t$, this would have indicated a change about a zero datum line.

By combining (1) and (2) and taking $\phi=0°$, we obtain an expression for an amplitude-modulated carrier-current.

$$i = A \sin \omega_c t\ (1+k_a \sin \omega_s t) \quad . \quad . \quad . \quad . \quad (3)$$

This formula indicates that the synthesis wave consists of a high frequency carrier, constant in frequency but which is being varied in amplitude in accordance with the signal wave about a datum displaced A from zero, Fig. 8 depicting the case of a sine-modulated carrier, 100% modulated.

Expression (3) can be expanded as follows :

$$i=A \sin \omega_c t + \frac{k_a A}{2} \cos (\omega_c-\omega_s)t - \frac{k_a A}{2} \cos (\omega_c+\omega_s)t \quad (4)$$

This indicates that this same modulated carrier can be considered to be built up of a spectrum of constant-amplitude, constant-frequency waves consisting of the original carrier-

frequency and two sets of high-frequency waves known as side bands, the amplitude of which will depend upon the value of k_a. Thus if k_a is unity each side band will consist of a single high-frequency wave of half the carrier amplitude as shown in Fig. 8.

The signal [1] itself, if not of pure sine form, may be analysed into a series of harmonic waves (by Fourier expansion) and can be shown diagrammatically as in Fig. 7a, the actual width of the band and the amplitudes of the different harmonics being dependent upon the character of the signal wave. Thus instead of the term $(\omega_s t)$ in equations 3, it is necessary to insert the expression for the complex wave, the effect of which is to produce a series of side-band waves, two for each of the component frequencies of the signal.

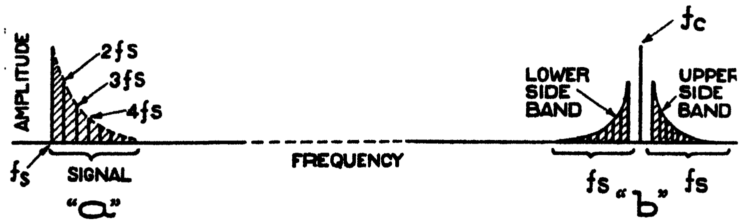


FIGURE 7.

The effect of modulating a high frequency carrier by a signal, therefore, may be considered as producing a high frequency spectrum which is obtained by moving the signal wave, depicted by Fig. 7a, along the frequency base by the amount of H.F. carrier. In addition to this a reversed image is created on the lower side of the carrier, as is shown in Fig. 7b, from which we can see that the high-frequency spectrum is twice the width of the signal band, and independent of the carrier frequency or percentage modulation.

This high-frequency spectrum consists of the following essential parts :

(1) The original carrier.

(2) A band of H.F. waves obtained by taking the sum of carrier and signal frequencies, called the upper side band.

(3) A band of H.F. waves obtained by taking the difference of the carrier and signal frequencies called the lower side band.

[1] It should be observed that the term signal is used in this chapter to denote the modulation and not a radiated signal complete with carrier.

The amplitudes of the side-band waves, however, depend upon the degree of modulation, and the power required to produce them represents the additional power necessary to effect modulation of the carrier.

Power Distribution in Modulated Carrier. It is of considerable interest to discuss the distribution of power in

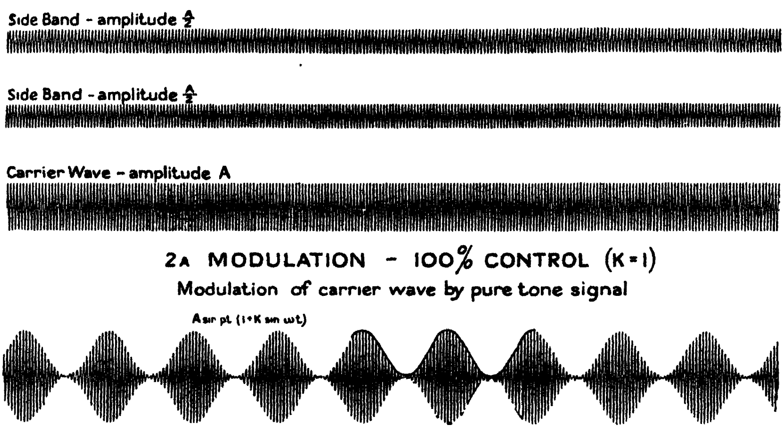


FIGURE 8.

such a built up wave and to observe the effects of phase displacements. The question is easily studied with the aid of a mechanical synthesis machine [1] and in the discussion which follows, actual examples of built up modulated waves taken by such a machine developed by one of the authors will be used to illustrate the various points.

Graphical examples of a sine modulated carrier with different percentage modulation are shown in Figs. 8, 9 and 10. Considering Fig. 8, for 100% modulation, the distribution of power will be proportional to the square of the individual waves, and it is seen that there is twice as much power in the carrier as there is in the two side bands together. If we have a carrier 50% modulated, as shown in Fig. 9, then the amplitude of each side band will be one-quarter the carrier

[1] Marconi Reviews, Nos. 9 and 10.

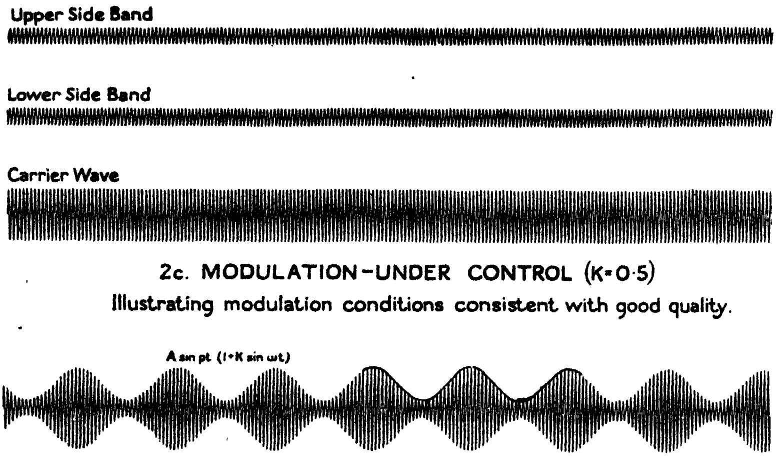


FIGURE 9.

amplitude, and hence the total side-band power is only one-eighth the carrier power.

If we modulate too deeply, side-band power is produced in too great a proportion and the result is a distorted wave, as shown in Fig. 10.

The power proportion of carrier to side bands is, of course, not a constant for a given percentage peak modulation, but varies with the shape of the signal wave, the more peaky the wave form the less power there is in the side bands for a given

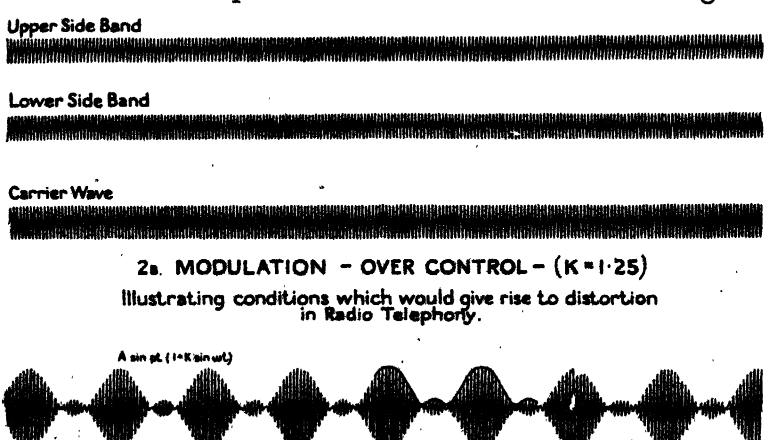


FIGURE 10.

value of peak amplitude. It really comes to a question of wave area, and examples in Figs. 11 and 12 show this. Fig. 11 shows a carrier modulated by a signal wave having a third harmonic with such a phase that the harmonic does not raise the peak of the wave but merely broadens it. Such a signal having one harmonic (compare envelope with Fig. 24) will have two side band waves each side of the carrier; one pair for the fundamental and one pair for the harmonic. Thus, for 100% modulation, a fundamental signal amplitude of half the carrier can still be retained without over modulating and the amplitudes of carrier, signal fundamental side bands, and harmonic side bands will then become as shown; that is, the fundamental side bands are half the amplitude of the carrier, and the side bands produced by the signal harmonic are one-sixth the amplitude of the carrier. The power distribution in such a wave will be, therefore, 1 in carrier to .556 in side bands, showing that because the wave is broader for the same peak value its power content is greater than in the sine case. Had the third harmonic been included such that it came in phase with the peak of the signal fundamental, thus

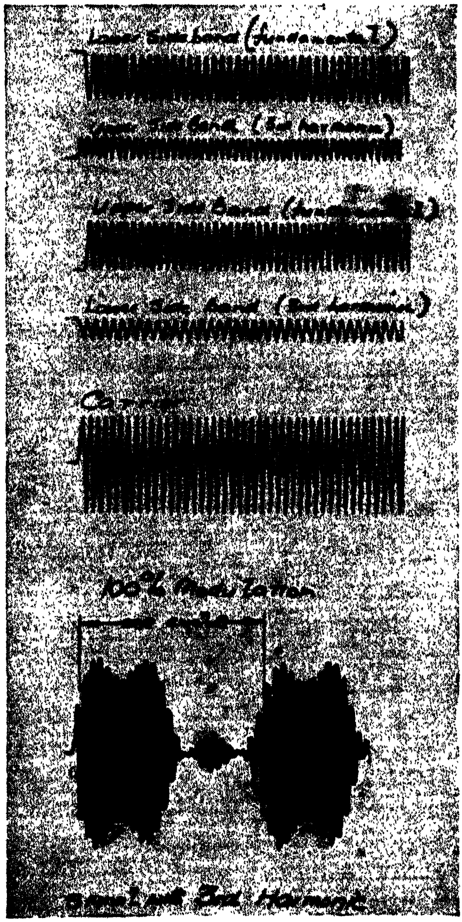


FIGURE 11.

making the envelope peaky, as shown in Fig. 12, with the same carrier it would be necessary to reduce the amplitude of the signal component to prevent over-modulation. In this case the power content of the two side bands becomes .312 of the power of the carrier for 100% modulation.

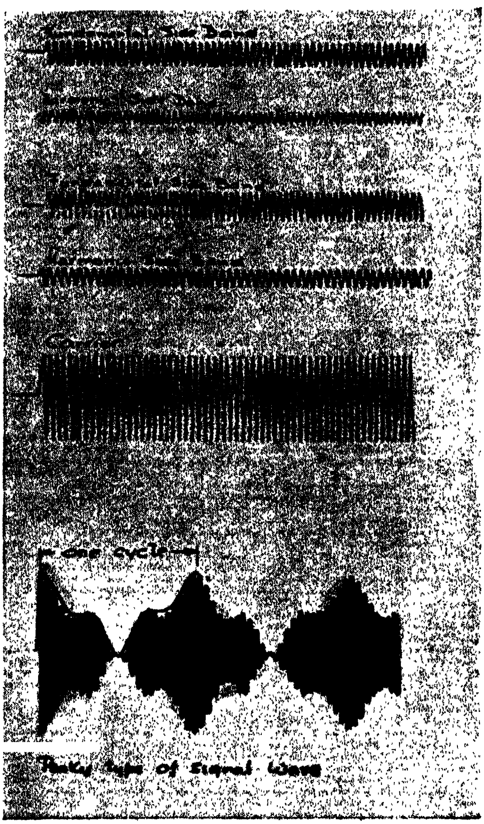


FIGURE 12.

The type of modulation containing the most power is a square wave, as it has the greatest area for a given maximum amplitude, and in this case the power in the side bands rises to equal that of the carrier, for 100 % modulation.

Thus, given the most favourable conditions, there will always be as much power expended in the carrier as there is in the side bands.

Vector Analysis of Amplitude Modulated Wave.

Although we have a spectrum of frequencies to deal with, ordinary vector analysis applied to an amplitude modulated wave is a simple conception because the side-band waves are so disposed about the central carrier-frequency that each pair combines to form a frequency equal to that of the carrier. Consider the case of a sine-modulated carrier. We have to think of the addition of three rotating vectors, a carrier-vector C and two side-band vectors S_1 and S_2 of appropriate length, as shown in Fig. 13. Since at all moments one side band is

gaining on the carrier as much as the other is losing, we can, for convenience, consider the carrier vector stationary and the side-band vector S_1 rotating at a uniform speed of f_s cycles per second in a clockwise direction and S_2 rotating at a uniform speed of f_s cycles per second in an anti-clockwise direction. The addition of these three vectors will obviously result in a vector stationary in position (of the carrier frequency), but varying in amplitude at the modulation frequency, the cyclic variation of amplitude being obtained from the instantaneous resultant side-band vector S_R and the carrier vector C. Since

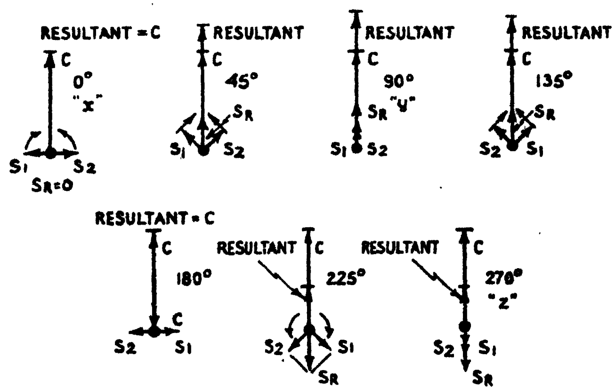


FIGURE 13.

this vector group will be rotating at constant speed it is clear that the resultant wave is one of constant frequency, varying in amplitude.

Such a method of visualising a modulated carrier provides a simple analysis, for the addition of any group of upper side-band vectors to its corresponding group of lower side-band vectors must always produce resultants which are exactly additive to or must be subtracted from the carrier directly.

Observe that the side-band vectors are rotating at constant angular velocity, once for every cycle of modulation and the angle traced out is equal to the proportion of modulation cycle completed, that is to say, since one cycle of modulation is represented by a complete revolution of each side-band vector, a quarter of a cycle of modulation is represented by angular positions of $\pm 90°$ of the vectors, half a cycle by $\pm 180°$ and so

on. If we change the modulation frequency we change the rate of rotation of the side-band vectors, but one cycle of modulation will still be represented by 360° angular change, and in consequence there is always the same simple direct relationship between the modulation cycle and the angle traced out.

By plotting the vector amplitude on a time base we can show the variation of modulation envelope and if the ratio of carrier to modulation frequency is known we can also show the varying carrier that builds up this modulation envelope. It is to be noted that as long as the vectors are rotating at constant angular velocity no change of frequency is involved.

Carrier - Suppression.

FIGURE 14a.

FIGURE 14b.

In order to effect economy of working, the suppression of the carrier from the transmitter has been suggested, and we will discuss such a case.

If we suppress the carrier from the synthesis wave, the wave radiated and received is the beat produced by the side bands alone, and for the case of a sine signal, the radiated and received wave will appear as shown in Fig. 14a, no matter what was the original percentage modulation.

If we trace out the signal component it will be seen that the datum line of the signal wave now coincides with the zero line of high frequency, thus conforming to the definition of a non-carrier system previously given. An interesting point to observe, shown in Fig. 14a, is the phase reversal occurring at every half cycle of modulating component, which is analogous to the phase reversal of current in the line case. If reception of this carrierless spectrum is considered it will be clear that detection will not give the original signal because the detector is unable to interpret the meaning of the phase reversal at the half cycle, and in consequence the signal received will be of double frequency and distorted, an exactly analogous case to the unpolarised telephone. To obtain the

original signal it is necessary to polarise the receiver with a high frequency carrier of the correct frequency and having a correct phase. The automatic phase reversal shown in Fig. 14a makes it clear how the addition of a polarising carrier at the receiver can build up the correct envelope shape. For the side bands, each being displaced on the frequency band an equal amount above and below the carrier frequency, when added to each other form a beat wave of the carrier frequency; thus if a carrier is added so that it is in phase with the one-half of the modulated wave it raises the envelope up; and because of the phase reversal, the carrier automatically assumes an antiphase condition with the second half cycle, the result being a corresponding reversal of envelope shape, and a correct reproduction of the original signal wave.

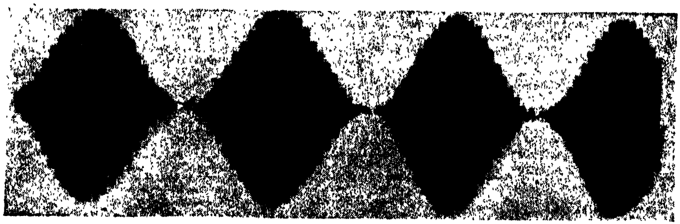
Fig. 14b shows the side-band spectrum for a signal with a third harmonic. Here again the datum line of the signal wave will be found to coincide with the zero line of high frequency and the phase reversal at the half cycle of signal is still in evidence.

It will be realised that it is extremely difficult to obtain two or more high frequencies sufficiently constant that they maintain the same phase relationship over a period of time (particularly on short waves), and we must therefore discuss the effect of phase shifts on the resulting envelope. It should be pointed out in passing that one can only talk about phase relationship between waves of different frequencies at some point of reference, and in the particular case under discussion this is the phase of carrier at the instant of time when the side-band waves are in phase opposition; such a point of reference in the case of a sine modulated wave indicates the commencement of the signal envelope.

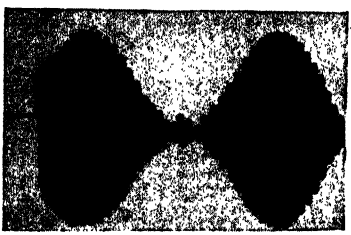
Effects of Phase Change in Re-introduced Carrier. Before considering a signal of complex shape it will be of interest to show the effect of carrier phase-shift on a sine-modulated wave with deep and shallow modulation, and the curves for this case are shown in Figs. 15, 16 and 17.

Fig. 15 shows that with deep modulation (100%) and a pure tone signal, considerable distortion occurs with very small phase shifts of carrier. Figs. 16 and 17 depict the case where the percentage modulation is 50% and 25% respectively.

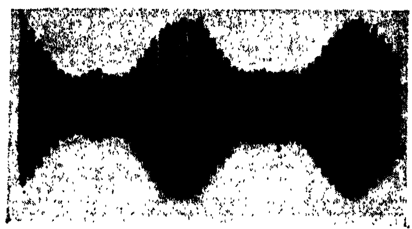
These curves show two remarkable features; first, a decrease in the effective modulation; secondly, no change in the position of modulating envelope on the time base but the introduction of distortion, but only as the carrier approaches quadrature. Thus considering the case of a signal (50% or



0° phase.



10° phase.



70° phase.

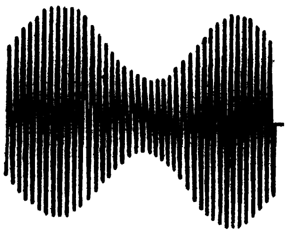
90° phase.

FIGURE 15.

25%) modulated, as the carrier phase is shifted from 0° to 90°, the depth of modulation progressively decreases, slowly at first, but rapidly as the phase shift approximates to quadrature until with a phase shift of exactly 90° the modulation has almost completely disappeared.

If the shift of phase is increased beyond 90° the modulation reappears with a reversed sense, and with a phase shift of

180° of carrier the signal resumes its original depth of modulation, but completely reversed. It can be imagined that for a signal envelope to go through a process of being turned inside out, as it were, necessitates a transition stage where the effective modulation is decreased as shown. With the pure tone signal

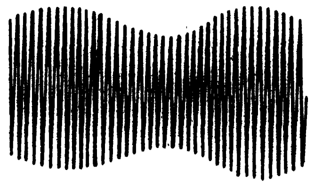


0° phase.

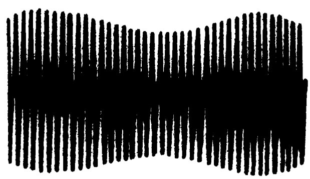
30° phase.

90° phase.

FIGURE 16.—50% Mod.



0° phase.



70° phase.

90° phase.

FIGURE 17.—25% Mod.

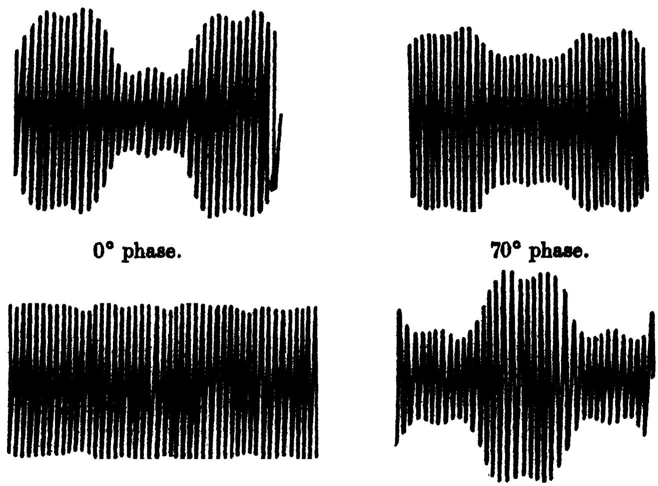
and deep modulation this is accompanied by distortion. It is important to observe that this demodulation [1] is not the only effect present, for a careful examination of the resulting wave form will show that the radiated wave is now not quite constant in frequency but varies between limits which increase as the apparent depth of modulation decreases. This will be referred to later.

Fig. 18 shows an example of a carrier wave 50% modulated

[1] The word demodulation is used to indicate a process which reduces modulation and not as an alternative term for detection.

by a signal having a strong third harmonic, the phase shift of carrier being indicated on the diagram. These figures show how little distortion appears as the carrier phase is changed, and it is seen that there is no relative phase shift of harmonic to signal fundamental.

That is to say, the effect of a shift of carrier phase for shallow modulation in ordinary telephony is to bring about a demodulation effect and introduce distortion. That is assuming such a wave be received by a system which responds to amplitude change, any small frequency variation passing unobserved.



0° phase. 70° phase.

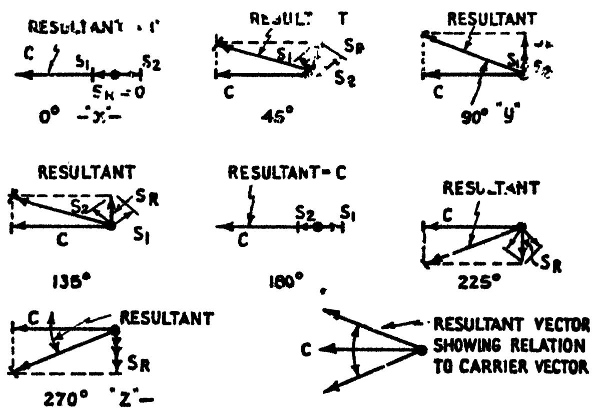
90° phase. 180° phase.

FIGURE 18.

Vector Analysis of Suppressed-Carrier System. Consideration of the vector analysis shows how these results come about.

If now we consider a carrier having a phase shift of say 90° it means that the sum of the side-band vectors must not be added directly to the carrier, but in quadrature, as shown in Fig. 19 ; because of this, if the carrier vector is large compared with the side-band vector (as it is for shallow modulation) the resultant amplitude of these two vectors added in quadrature will always be approximately the same, for a small vector added in quadrature to a large vector gives a resultant of almost the same amplitude as the larger vector, as shown in Fig. 19.

Considering the rotation of this group of vectors, our carrier vector is the reference point, and it is clear that the vector resulting from the carrier and the side-bands no longer maintains a constant position relative to the carrier, but it is at one time leading it, and at one time lagging behind it; in fact there is a cyclic variation at the modulation rate, and we therefore come to the conclusion that the resulting wave is no longer constant in frequency but varies cyclicly at the modulation frequency, this frequency variation being additional to the small cyclic change of amplitude.



THE RESULTANT VECTOR SWINGS EITHER SIDE OF CARRIER BETWEEN LIMITS DETERMINED BY THE PERCENTAGE MODULATION THIS CONSTITUTING A FREQUENCY VARIATION.

FIGURE 19.

Phase-shifting the carrier of an ordinary amplitude-modulated wave has the effect, therefore, of reducing the amplitude modulation and introducing a frequency modulation; and with a carrier phase-shift approaching quadrature, the wave may become almost demodulated, although a change of frequency will have been introduced which may become fairly considerable.

If a carrier is introduced of a different frequency it will beat with the side-band spectrum and there will be transient conditions only, where the envelope is reproduced correctly.

The conclusions one arrives at are, therefore, that to prevent signal demodulation we must hold the carrier phase to within

a few degrees per cycle, and since this is an impossibility with present day technique, because of insufficient frequency stability, a suppressed carrier system is not a practicable proposition.

Single Side-Band Working. To overcome the difficulty of working a suppressed-carrier system a compromise has been effected by adopting what is known as a "single side-band working," and this system is of considerable interest, for it

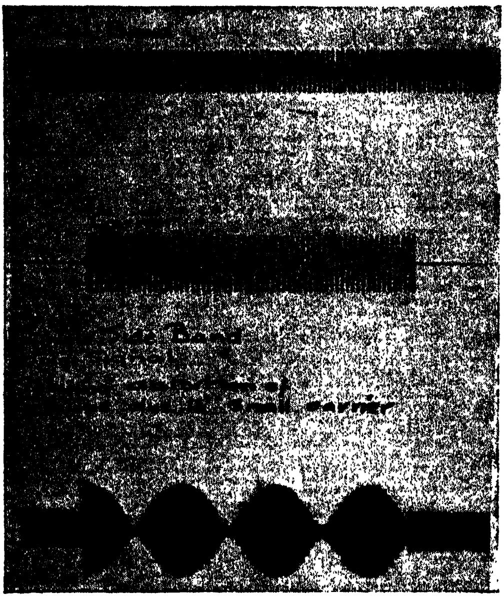


FIGURE 20.

reduces the frequency spectrum used. It should be pointed out, however, that single side-band working has no power economy over full side-band working with a suppressed carrier, and it has its limitations, as will be seen.

Consideration of the original high-frequency spectrum shows that either side-band contains the signal components. Single side-band working, therefore, consists of suppressing at the transmitter not only the carrier but one side band as well, and thus one group of waves is transmitted having the frequency of carrier plus (or minus) the signal component frequencies. Suppression of the carrier can be carried out by a balanced circuit, as described in Chapter IX, the side-bands being

separated by any convenient form of filter circuit. As in the case of suppressed-carrier working, to reproduce the signal at the receiver it is necessary to add a polarising wave of the same frequency as that suppressed, since detection of the side-bands alone does not result in the extraction of the signal component.

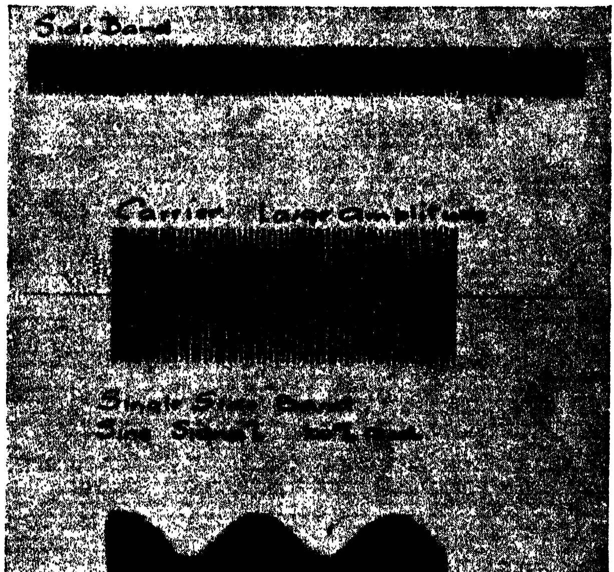


FIGURE 21.

The theory of action is that because the carrier differs from the waves of the side-band by amount of the signal components, the beating of the group with a similar carrier produces a synthesis wave whose envelope is the signal.

Thus in the case of a sine signal of frequency f_s, modulating a carrier of frequency f_c, two side-band waves are produced (f_c+f_s) and (f_c-f_s). Thus, if the upper side band is being used it will consist of a frequency (f_c+f_s). The addition of a carrier frequency f_c at the receiver before detection

produces a wave whose beat envelope will have a frequency of $(f_c+f_s)-f_c$, namely f_s, the original signal frequency.

This simple arithmetic, however, is somewhat misleading, and it will be found that although f_s does appear as the chief frequency in the detector output, the original envelope shape is not reproduced, although under certain conditions

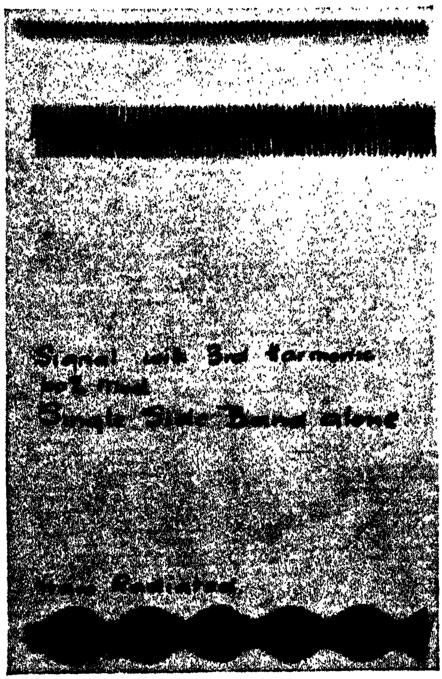


FIGURE 22.

the distortion can be reduced to a small amount. Consider the above case. Single side-band working for a sine signal is a transmission of a single wave differing in frequency from the carrier by the amount of the signal frequency. Addition of a local carrier produces an envelope of the original signal frequency, as shown in Figs. 20 and 21, but from observation this envelope is certainly not a sine wave. If the carrier added is large in amplitude compared with the side-band, the envelope appears to approach more to a sine shape, as is shown in Fig. 21 (compare with Fig. 20), and this indicates that for single side-band

work a very large local carrier is necessary, if good reproduction is required

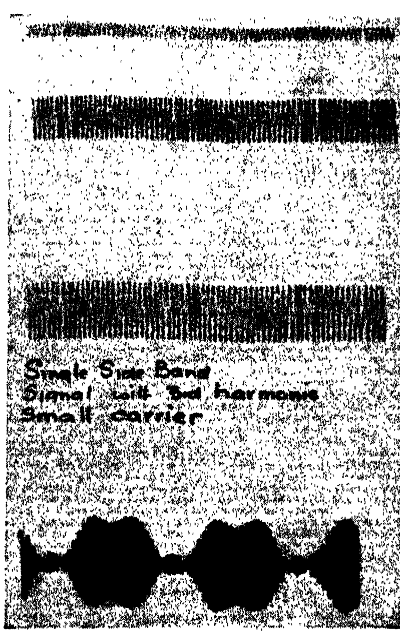
Effect of Phase Changes in Single Side-band System. It is of interest now to discuss phase and frequency changes of the added carrier. Considering the sine-signal case, since the side-band consists of one wave only, it is obvious that the same envelope will be produced whatever phase is considered, although of course the phase of envelope is not a constant. If the signal is not a pure sine wave the different frequencies comprising it will all get different phase shifts, as will be shown.

Consider a signal having a third harmonic. The original modulated carrier wave for such a signal produced by the transmitter is as shown in Fig. 11. If we suppress the carrier and one side band, what is radiated and received will be one side band only. If we assume the "lower" is used, Fig. 22 shows what this radiated wave comprises, from which we can see that the envelope has no obvious relationship to the signal.

If carefully examined it is found that the envelope is actually the difference frequencies of the wave making up the signal. Thus in the present case of a signal with frequencies 1 and 3, the difference is 2, shown by the envelope.

With a pure tone, since there is only one frequency, the difference frequency is 0, that is the single side-band wave for a pure tone is a continuous wave.

As stated previously, to be able to obtain the correct intelligence it is necessary to add a carrier to this side-band wave before detection, and we will consider



(a) 0° phase.

FIGURE 23.

first the introduction of a carrier of correct frequency but having correct or incorrect phase.

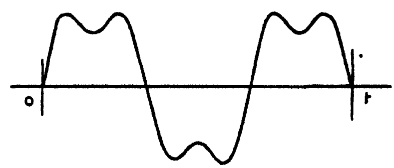
Fig. 23a shows a carrier introduced with correct phase, Fig. 23b with a carrier 45° phase shift, and Fig. 23c with 90° phase shift. Had the phase shift been 180° the envelope would have been completely reversed in sense. At first sight these envelopes appear quite different, but actually they all contain the same audio frequency components with different relative phase shifts, and we must consider what effect this has.

(b) 45° phase.

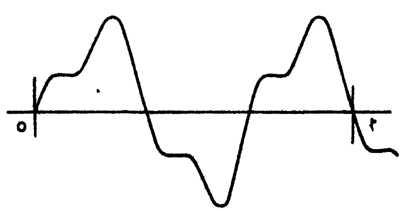
(c) 90° phase.

FIGURE 23.

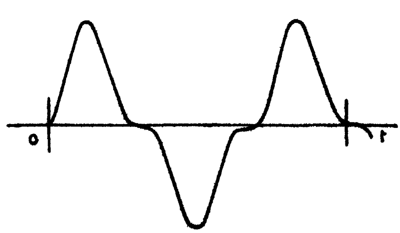
In Fig. 24 is shown the synthesis wave of a signal with third harmonic, the latter having different phase shifts as indicated. If these wave shapes are compared with the modulated envelopes of Fig. 23 it is clear that the latter can always be identified with a signal having a third harmonic. But since the envelopes are different it means the phase shift of the harmonics is different from that of the signal fundamental. Considering the case of the carrier introduced with a 90° phase shift, it appears that whereas the fundamental of the wave is shifted ½ cycle along the time base the third



3RD HARMONIC IN PHASE



3RD HARMONIC IN 90° PHASE SHIFT



3RD HARMONIC 180° PHASE SHIFT

FIGURE 24.

harmonic is shifted ½ of one quarter cycle of the fundamental signal frequency, i.e. into anti-phase with the signal fundamental. This means that with a wave rich in harmonics, not only will the phase of the fundamental be shifted but the phase of each individual harmonic shifted a different amount, and if one considers the general case it is observed that the phase shift of the different component frequencies is directly proportional to those harmonic ratios.

Because the shape of the resulting envelope depends upon the relative phases of the different components of the signal it follows that in single side-band working, the envelope shape alters considerably with a phase shift of carrier. If, however, such a wave is being received aurally, the intelligence conveyed by the ear to the brain will be exactly the same whatever envelope shape is made by the combination, for the ear is only capable of interpreting frequencies and amplitudes, and not phases. It interprets frequencies accurately, amplitudes indifferently, and phases not at all, and this statement is true for all waves which are not transient in character.

If the added carrier is altered in frequency, the pitch of the fundamental will alter, and since the difference frequency to the harmonics changes, the signal harmonic frequencies will alter relative to the signal fundamental, but it should be observed that because the signal fundamental has the smallest difference of frequency to the carrier, a small change of carrier frequency makes a larger percentage change in the fundamental of the signal than it does in the harmonics. Now if one considers an aural signal, the fundamental conveys mostly pitch and the harmonics mostly character, and because the harmonics are changed comparatively slowly compared with the fundamental, one can allow quite a remarkable amount of carrier frequency change before a speech signal becomes unintelligible, the most marked alteration being the rapid change of voice pitch. The above statements are true for speech but not for the transmission of music, for in the latter case much of the pleasure (or otherwise) lies in the combination tones of the various frequencies, and if these are upset the result is discord.

The general inference is, then, that single side-band working offers a wide field of development for certain classes of traffic,

but that field will be limited to systems employing aural reception, unless distortion of envelope can be permitted, frequency characteristic of circuits improved or synchronising systems adopted.

Actually no intelligence can be conveyed by a single side-band system unless :

(*a*) The intelligence being conveyed is known to the receiving operator, or :

(*b*) The exact carrier frequency originally suppressed by the transmitter is known at the receiving station.

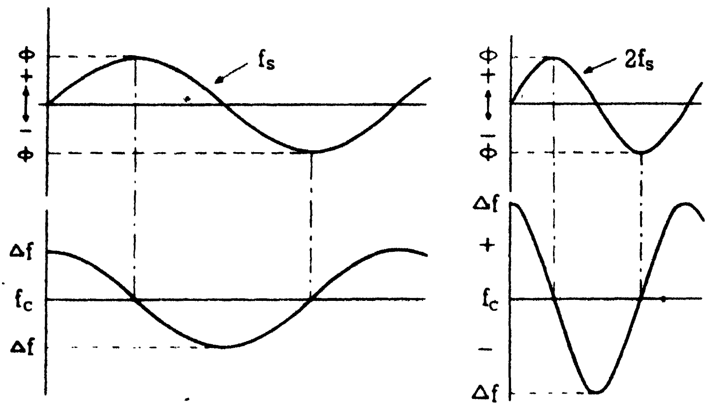


FIGURE 25.

In the case of telephony, for instance, the receiving operator knows that he has to receive intelligible speech and in consequence automatically adjusts the system until this is obtained. But suppose the intelligence required is a pure tone, then since the single side band is now but a single frequency of constant amplitude, the receiver can in no way collect the original intelligence accurately.

Phase and Frequency Modulation. As indicated in a previous section of this chapter, alternative methods of modulation involve phase or frequency changes instead of amplitude changes. In phase modulation, the signal produces a change in the phase of the carrier $\Delta\phi$ which is proportional to the instantaneous amplitude of the signal irrespective of the signal frequency, as shown in Fig. 25 ; whereas with frequency modulation, the signal produces a frequency deviation f Δ

proportional to the instantaneous amplitude of the signal irrespective of the signal frequency, as shown in Fig. 26.

If we consider the application of a sinusoidal signal, both systems are essentially the same in effect, because a sinusoidal phase displacement represents a frequency deviation proportional to the rate of change of slope of the phase-time curve, and this is still sinusoidal as indicated in Fig. 25, which shows the phase displacement time curve and the resultant frequency deviation time curve beneath. It will be observed from Fig. 25, where curves for two different signal frequencies are given, that at points of maximum phase displacement the rate of change of phase is negligible and in consequence the frequency

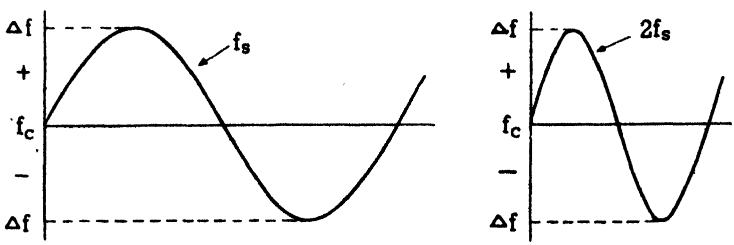


FIGURE 26.

is that of the unmodulated carrier; whereas at points where the phase displacement is changing most rapidly, the frequency deviates from the carrier by an amount which is proportional to this rate of change of signal, the higher frequency giving a greater deviation frequency. Thus, with a phase modulated carrier, although the phase displacement is a constant for a given amplitude of signal, the rate of change of phase and in consequence the deviation frequency increases with the signal frequency, the ratio $\frac{\Delta f}{f_s}$ remaining a constant.

Although there is a similarity between a phase and a frequency modulated wave for an applied sinusoidal wave, this is only because the differential of a sine wave is still sinusoidal in form; for other types of signals there is a distinct difference. For instance, consider the application of a signal having a square wave-form. For the case of phase modulation, the wave remains constant in frequency f_c, but suddenly changes phase every half cycle of modulation, whereas in the case of a

frequency-modulated wave the frequency of the carrier will suddenly change every half-cycle between limits of $(f_c+\Delta f)$ and $(f_c-\Delta f)$.

Phase Modulation. Consider a carrier wave whose amplitude A and frequency f_c remain constant, but whose phase is varied sinusoidally about a mean value ϕ by an amount $\pm k_p\phi$ where $(k_p\times 100)$ is the percentage modulation. By following reasoning similar to the amplitude modulated case we have:

$$i=A\sin\left\{\omega_c t+\phi\,(1+k_p\sin\omega_s t)\right\}$$

$$=A\sin\left(\omega_c t+\phi+k_p\phi\sin\omega_s t\right). \quad . \quad . \quad . \quad (5)$$

Let $m_p=k_p\phi$ and assume the carrier phase-angle is 0°. Then it can be shown that:

$$i=A\left[J_o(m_p)\sin\omega_c t+J_1(m_p)\left\{\sin(\omega_c+\omega_s)t-\sin(\omega_c-\omega_s)t\right\}\right.$$

$$+J_2(m_p)\left\{\sin(\omega_c+2\omega_s)t+\sin(\omega_c-2\omega_s)t\right\}$$

$$+J_n(m_p)\left\{\sin(\omega_c+n\omega_s)t-\sin(\omega_c-\omega_s)t\right\} \quad . \quad (6)$$

Where $J(m_p)$ $J_1(m_p)$ are Bessel functions with argument m_p. From tables of Bessel functions the values of the coefficients may be found for any value of percentage modulation. Examination of equation 6 and reference to the amplitude coefficients would show that the modulated wave would comprise a carrier wave and two groups of side-bands of frequencies

$$f_c\pm f_s,\ f_c\pm 2f_s,\ f_c\pm 3f_s \quad . \quad . \quad . \quad . \quad . \quad f_c\pm nf_s.$$

Unlike the amplitude-modulated wave, however, each modulation frequency produces an infinite spectrum of side-band pairs, instead of a single pair; moreover the carrier amplitude does not remain unchanged, but has an amplitude which depends upon the percentage modulation, and the phase of side-bands relative to the carrier is not zero (as previously defined), but that of each pair alternates between zero and quadrature. If reference is made to a previous example of quadrature phase of a single pair of side-bands (see page 38), it was seen that the resultant had not only a frequency variation but a small amplitude variation as well; the introduction of

this infinite series with alternating phase is to level out this amplitude difference.

Fig. 27 shows (left hand) the frequency spectrum (half) for a phase modulated carrier for an applied signal frequency of 5 kc/s and for different values of k_p and for a max. phase displacement of $\phi=10$ radians, in which side-bands less than $\frac{1}{1000}$th of the unmodulated carrier have been neglected. These

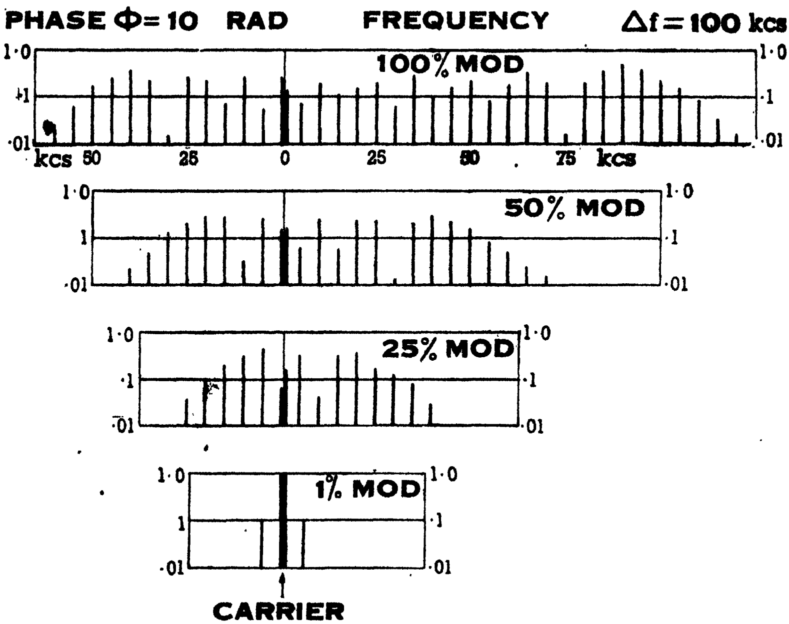


FIGURE 27.

figures show that the width of the spectrum is not dependent upon the signal frequency necessarily, as with amplitude modulation, but also upon the equivalent deviation frequency produced, which in turn is dependent upon the phase angle and percentage modulation. Although there is no simple relationship between band-width and the ratio of signal to deviation frequency, in general it is determined by whichever is the greater, and where one is much greater than the other, the band width is roughly twice the greater frequency. For instance, in the example given of $\phi=10$ radians for 100% modulation, a deviation frequency Δf of 50 kc/s is produced for

a 5 kc/s signal, and the band width is in this case rather greater than twice Δf; whereas with a low percentage-modulation (1%) the deviation is only one-tenth the signal frequency, and in this case the band width is approximately twice the signal frequency.

Frequency Modulation. Following similar reasoning for the phase modulated case, it can be shown that a frequency-modulated current wave may be expressed:

$$i=A\left[J_0(m_f)\sin\omega_c t+J_1(m_f)\left\{\sin(\omega_c+\omega_s)t-\sin(\omega_c-\omega_s)t\right\}\right.$$
$$+J_2(m_f)\left\{\sin(\omega_c+2\omega_s)t+\sin(\omega_c-2\omega_s)t\right\}$$
$$\left.+J_n(m_f)\left\{\sin(\omega_c+n\omega_s)t-\sin(\omega_c-n\omega_f)t\right\}\right] \quad . \quad (7)$$

where $m_f = k\,\dfrac{\Delta f}{f_s}$

and Δf is the deviation frequency for 100% modulation. Comparing this with expression (6) it is clear the resultant spectrum for a frequency-modulated wave will be very similar in form to that for a phase-modulated wave, the only difference being in the value of the coefficients m_p and m_f. At the signal frequency for which these coefficients are equal the two spectra will be identical, but any change of frequency would produce a difference. We can say that as with phase modulation, the band-width is in general determined by whether Δf or f_s is the greater, and where one frequency is much greater than the other, the band-width is approximately twice the greater frequency, the frequency spectrum for a frequency modulated carrier being shown in Fig. 27 (right) for a signal of 5 kc/s, and Δf = 100 kc/s.

Vector Analysis of Phase and Frequency Modulation. With amplitude modulation we were interested in a synthesis wave of constant frequency and varying amplitude. The vector representation of such a wave is explicit and simple because the length of the vector shows directly the instantaneous amplitude, whilst the angular position from a given datum shows the phase relative to that datum. Because we are dealing with a constant frequency, the constant angular rotation of the vector scarcely enters into the argument. The use of this same vector convention to express a phase or

frequency-modulated carrier becomes more difficult, because we have a carrier whose frequency is varied. Thus instead of a vector rotating at constant speed but varying in amplitude per modulation period, phase or frequency modulation will be represented by a constant amplitude vector having a cyclic variation of angular position from a given datum, and a variation of angular velocity per modulation period, a rather more difficult conception.

It may be rather easier to explain if we consider first a carrier f_c (say 100 c/s) modulated by a square-wave signal

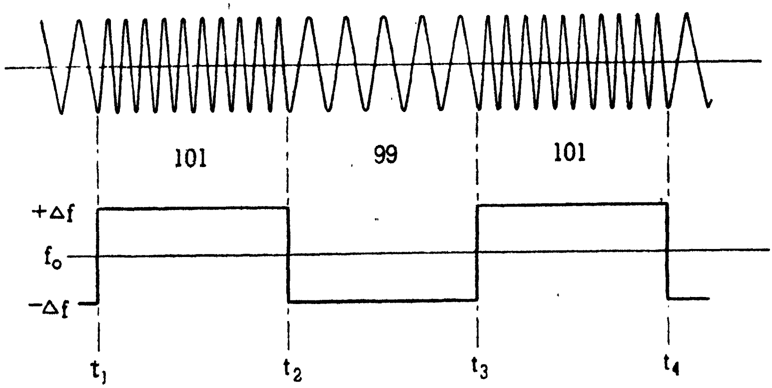


FIGURE 28.

to a maximum deviation frequency Δf of say 1 c/s, at a signal frequency f_s, for which we will take different values as mentioned later. Such a modulated carrier would be represented as shown in Fig. 28, where modulation of the carrier is carried out by increasing the frequency to 101 c/s from time t_1 to t_2, reducing it suddenly to 99 c/s, shown from t_2 to t_3, the rate at which this periodic change is made being of course the modulation period.

As we are assuming the changes from 99 to 101 and back are made instantaneously, there are transient conditions only at t_1, t_2, t_3, etc., where the frequency passes through the carrier value of 100 c/s.

Vectorially, such a frequency-modulated wave will be represented by a vector which rotates at 101 c/s for a period t_1 to t_2, then suddenly slows up to 99 c/s for the period t_2 to

Note.—Frequency in Fig. 28 not to scale.

t_3 and so on. Or, if we adopt the previous convention, namely, to represent the carrier by a stationary vector, then for the period from t_1 to t_2 the vector would rotate in an anti-clockwise direction at a constant speed, and from t_2 to t_3 in a clockwise direction at a constant speed of 1 c/s, as indicated in Fig. 29. The transient carrier condition will be shown by the stationary vector at the ends of the arc of travel, the uniform speed of vector in either direction being dependent upon the deviation frequency, in this case 1 c/s.

The total arc of travel of vector, however, is not dependent upon the deviation frequency alone but upon the modulation cycle, because the longer the time taken to carry out this cycle (or the less the modulation frequency) the longer time will the vector have to rotate in either direction. For instance, if the total time of the modulation cycle is one second the vector will have half a second to rotate in either direction, and since it is travelling at a uniform speed of 1 c/s, the vector will trace out an arc of 180°. If we increased the rate of modulation (leaving the deviation frequency the same) the arc of the vector will be reduced. Thus with a modulation frequency of 10 c/s, the arc of travel will be only 18°.

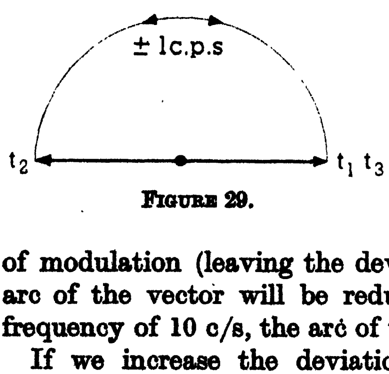


FIGURE 29.

If we increase the deviation frequency, we increase *the velocity at which the vector is travelling*, and in consequence the length of arc will be increased proportionally. Thus if the deviation frequency is increased to 10 c/s, and we perform the operation once a second, the vector will have time to travel 5 cycles both in an anti-clockwise and a clockwise direction. Increasing the modulation frequency to 10 c/s the arc of travel will be reduced to 1/10th of this, and thus is again brought to 180°.

Thus the arc of travel of vector is dependent upon the ratio of $\frac{\Delta f}{f_s}$ irrespective of the actual frequencies involved and irrespective of the carrier frequency. It might be observed here that with a frequency-modulated carrier, since the deviation frequency is constant for all signal frequencies this ratio

$\frac{\Delta f}{f_s}$ is not constant but decreases with increase of signal frequency, and thus the arc of travel will decrease with increase of signal frequency. But with a phase modulated carrier, since the ratio $\frac{\Delta f}{f_s}$ is a constant, the arc of vector travel is a constant for all signal frequencies, but the velocity of vector increases proportionally as the signal frequency is increased.

If now we consider the application of a sinusoidal signal, the change of deviation frequency is not made instantaneously at the points t_1, t_2, t_3, but sinusoidally throughout, and this

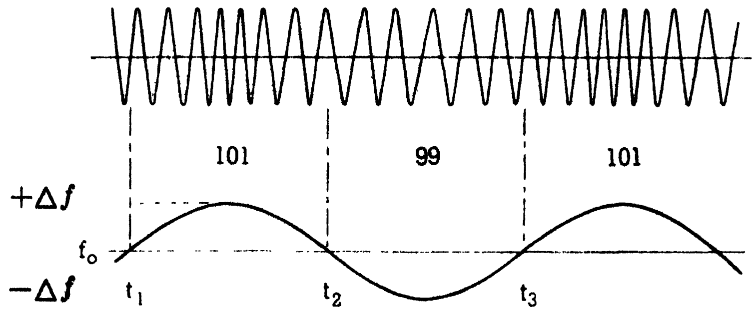


FIGURE 30.

will further modify the length of travel of the vector representing the cycle of events. For instance, consider the same carrier of 100 c/s deviated to a maximum frequency of $\Delta f = 1$ c/s as shown pictorially in Fig. 30, the frequency reaching its maximum and minimum values at points half-way between the times t_1, t_2, t_3, etc. The difference vectorially will be that the vector representing this changing frequency will not swing at a uniform velocity each way but with an angular velocity which varies sinusoidally throughout its swing. Thus at the ends of the swing the velocity is zero, representing the carrier frequency, and it gradually increases from zero to a maximum velocity representing the maximum deviation frequency at a point midway between the extremities of arc. Thus taking the previous example if the deviation frequency is 1 c/s and the rate at which the modulation is made is also 1 c/s, this will mean that the vector will have half a second to travel in either

direction, but since its velocity is not uniform (as with a square modulation signal), but only reaches the velocity of 1 c/s at the centre of swing, the vector cannot travel 180°, but only $\frac{2}{\pi} \times 180°$. As before, the arc of travel will also be dependent upon the ratio $\frac{\Delta f}{f_s}$.

We must of course clearly understand that this vector only relates the periodic change of two functions and to find the frequency at any point of the cycle, we must know the actual figures involved.

Note that such a swinging vector will equally well represent either a phase or a frequency modulated wave. But the phase angle will be directly correlated to the mid position of the vector, which will be the carrier position for a phase modulated wave, phase angles being traced out either side of this mid position. The deviation frequency resulting from the modulation of the phase will be shown as explained by the velocity of the vector, and as was seen with sine modulation the velocity is in quadrature with the phase angle.

Resume of Phase and Frequency Modulation. In general phase and frequency modulation systems are classed under the headings "narrow" and "wide," that is whether the deviation frequency produced is small or large compared with the signal frequencies.

There are two principal advantages which phase and frequency modulated systems have over an amplitude modulated system. The first is that the transmitter can be operated under more favourable conditions. If reference is made to Chapter X, page 286, it will be seen that a valve transmitter when delivering its peak output operates at its highest efficiency. Now it is clear with an amplitude modulated carrier the energy surges from high to low levels each modulation cycle and in consequence the average efficiency must be less than if the transmitter was maintained at peak output throughout the cycle. With frequency and phase modulation, because the amplitude remains constant throughout the modulation cycle, we can operate the transmitter at its high efficiency condition throughout.

The second advantage is the increased signal/noise ratio that can be obtained, particularly with a wide-deviation modulation. This is due to the fact that the noise is limited to a band width within audio range of the carrier, whereas the modulation spectrum is caused by a very much wider band of frequencies. For example, with a Δf of 100 kc/s. as shown in Fig. 27, although the power is carried by frequencies spread over the whole of this frequency range, the noise contribution is confined to a band width dependent upon the maximum audio frequency, e.g. 5 kc/s.

It would be out of place to discuss the advantages and disadvantages of the various systems as they concern the entertainment field principally and the future balance of the types adopted will be governed largely by geographic, political and economic considerations. It is sufficient to state that amplitude modulation is used in 99% of commercial apparatus; that frequency modulation is used to a small extent in telegraph services (see page 428), and wide-deviation phase and frequency modulation systems are being developed to produce supplementary broadcast services of high quality but limited range on u-s. waves.

SELECTED REFERENCES

Modulation in Radio Telephony. Heising. *P.I.R.E.*, Vol. 9; August, 1921.

Notes on Theory of Modulation. Carson. *P.I.R.E.*, Vol. 10; February, 1922.

Relations of Carrier and Sidebands in Radio Transmission. Hartley. *P.I.R.E.*, Vol. 11; February, 1923.

A Mechanical Synthesis Machine. Ladner. *Mar. Rev.*; June, July, August, 1929.

Frequency Modulation. Van der Pol. *P.I.R.E.*, Vol. 18; July, 1930.

Amplitude, Phase and Frequency Modulation. Roder. *P.I.R.E.*, Vol. 19; December, 1931.

Interference in Relation to Amplitude, Phase, and Frequency Modulated Systems. O. E. Keall. W.E. January, 1941.

CHAPTER IV

THE PROPAGATION OF SHORT AND ULTRA-SHORT WAVES AT SHORT DISTANCES

In this and the following chapter the very involved subject of wireless wave propagation will be treated only in a descriptive fashion. For the sake of completeness, it will be necessary to refer briefly to the longer waves but the main discussion will be concerned with short and ultra-short waves.

It is evidently very desirable that it should be possible to predict the signal-strength which will be obtainable at a distance from a transmitter of specified power and frequency, and using a given aerial system, etc., but the acquisition of sufficient data to do this is one of the most difficult problems in wireless engineering. Theoretical investigations require advanced mathematics and numerical solutions usually depend upon constants which must be determined experimentally. The measurements to determine the constants or to confirm the theories are difficult. Only simplified conditions (as, for example, the assumption of a simplified ionosphere and a smooth spherical earth of uniform conductivity and dielectric constant) are amenable to theoretical treatment in general.

Wireless telegraphy had been a commercial proposition for some time before any such measurements had been made, but progress in recent years has been rapid, due to the pains-taking research of numerous workers in many countries. As a result of the information collected, a Committee of the C.C.I.R. were able in 1939 to draw up a comprehensive report[15] giving quantitative curves and data covering the propagation, under average conditions, of wireless waves of all wave-lengths in use. This certainly does not mean, however, that finality has been reached in the study of wireless propagation, inasmuch as much more information is still required to clear

up a number of doubtful points and explain the numerous results which occur periodically and do not conform to type.

In fact the more statistical evidence that is accumulated the more one realises that the behaviour of the ionosphere is like that of the weather. Whilst one can give guiding rules as to its probable behaviour at a certain place, date, and time, the vagaries of the ionosphere will be certain to produce many incalculable results from time to time.

The study of wireless waves involves the behaviour of surface waves over the earth, and propagation through the earth's surrounding atmosphere. In short distance communication, using short and ultra-short waves, we are chiefly concerned with the former, and it would appear desirable to commence with a discussion of short distance propagation.

The Hertzian Dipole. The simplest arrangement to consider as a source of radiation is a straight conductor, very short compared with the wavelength produced and carrying a uniform alternating current. The radiation from practical aerials can be calculated by considering the aerial to be composed of a number of such elements.

The equation for the electric field at a distance from the dipole and on a plane perpendicular to the conductor is given by

$$E = 377 \frac{I h}{\lambda r} \text{ volts per metre}$$

$$\left.\begin{array}{ll} \text{where} & 2h \text{ is the height of the complete dipole} \\ & \lambda \text{ is the wavelength} \\ & r \text{ is the distance} \end{array}\right\} \text{in metres.}$$

$\quad$ and $\quad I$ is the current in amperes

It will be evident from the shape of a dipole, that if it is vertical, the radiation will be the same in all directions in a horizontal plane, and this radiation can, therefore, be illustrated by a polar diagram as in Fig. 31a, where the radius represents the field strength which would be measured if a suitable apparatus was carried round in a circle, having the dipole as its centre. A consideration of the field distribution in a zenithal plane shows that there will be no radiation vertically, maximum radiation horizontally, and for other zenithal angles it is reduced, following a cosine law, as shown in Fig. 31b. The

distribution in all directions is therefore represented by the three-dimensional figure sketched in Fig. 31c.

It will be observed from the formula that the field strength falls away as the first power of the distance, this reduction of signal strength simply being due to the spreading of the wave.

At points really close to the dipole, i.e. less than one wavelength, there are also the ordinary electric and magnetic fields

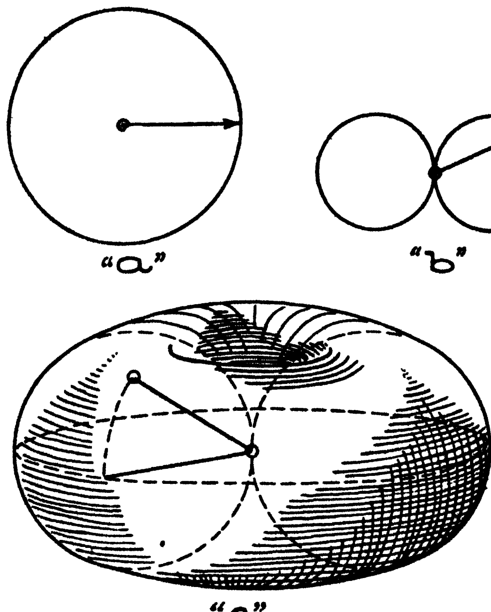


FIGURE 31.

to be considered in addition to the radiated fields. As, however, these induction fields are inversely proportioned to the square of the distance, they quickly become negligible and need not be considered in propagation problems generally, but they are important in problems connected with the behaviour of a number of radiators near to one another. It should be mentioned that the induction fields are in time quadrature with each other, whereas the radiation fields are in time phase.

We now require to know what power is being radiated away from the dipole, and as in the ideal case there is no loss in propagation, we can consider the energy flowing through a

sphere at a considerable distance from the dipole. Our equations give the field strength at every point in this sphere and from this we can determine the energy flowing through a unit area, so that summing over the surface we obtain the total power radiated. The result obtained by this method is:

$$W = \frac{320\ \pi^2\ h^2\ I^2}{\lambda^2}\ \text{watts.}$$

Since W is proportional to I^2, other things remaining constant, this power may be considered as used up in a fictitious resistance R_a, such that:

$$W = I^2\ R_a$$

and it will be seen that the radiation resistance of a dipole is given by:

$$R_a = \frac{320\ \pi^2\ h^2}{\lambda^2}$$

The modification necessary to adjust the calculations for the ideal dipole to fit ordinary aerials will be discussed later in Chapter VII.

The wave at a sufficient distance from such a dipole will be plane polarised, the electric and magnetic fields being mutually perpendicular to each other, and to the direction of propagation. In wireless engineering, the plane of polarisation is stated with respect to the electric field, and with reference to the earth's plane. Thus in a vertically-polarised wave the electric field is in a plane perpendicular to the earth's tangent plane.

If a perfectly-conducting, horizontal, plane sheet is passed through the centre of the vertical dipole, the fields should not be disturbed since the electric field is everywhere perpendicular to the plane. We deduce, therefore, that if we set up a half-dipole on a perfectly conducting plane, the field in the space above the plane is the same as that produced by a complete dipole in space. Another way of arriving at the same result is to suppose that the half-dipole has produced an "image" beneath the plane and that the current in the image is the same in magnitude and phase as that in the actual half-dipole.

Conductivity and Dielectric Constant of the Earth's Surface. In many cases the propagation of wireless waves over the earth's surface is much affected by the values of

conductivity (σ) and dielectric constant (κ) and the assumption of perfect conductivity made above would lead to wrong conclusions being reached. The values of σ and κ naturally vary considerably for different kinds of soil and there are also large variations with weather conditions at the same site. The soil at different depths will also have different properties and it is, therefore, not an easy matter to state an effective value for use in a wave-propagation problem.

Measurements undertaken by the Radio Research Board,[4, 5] showed very large variations in σ with moisture content. Thus for one sample of loam σ was 10^5 electrostatic units ($\rho = 9 \times 10^6$ ohms per cm. cube) when the moisture content was about 1% and $1{\cdot}5 \times 10^8$ (6000 ohms) when it was 25%, these figures being taken at a frequency of 1,200 kc/s. For the same sample, at the same frequency, κ varied from 3 for a 1% moisture content to 37 for a 25% content.

The conductivity increases with frequency whilst the dielectric constant decreases, the same sample as previously mentioned, when measured at 10 Mc/s, and at a moisture content of 25%, gave values of $\sigma = 2 \times 10^8$ and $\kappa = 30$.

Reflection at the Earth's Surface. We have already seen that if a vertical half-dipole is erected on the surface of a perfectly conducting plane, the field at a distance can be considered as due to a complete dipole. This is a particular case of the "image" theory by which the effect of reflection is taken account of by assuming an image carrying a current $I\ (A\,|\underline{\phi}\,)$, where I is the current in the actual dipole and $(A\,|\underline{\phi}\,)$ depends upon the type of reflection and in this case is $A = 1$, $\phi = 0°$.

The theory of reflection has been completely worked out, so that if σ and κ are known, the reflection coefficient applicable to a wave polarised either in the place of incidence or perpendicular thereto and of any frequency and incident at any angle can be determined. The principal practical difficulty is to know what values of σ and κ to assume, especially if there are changes in the nature of the soil just below the surface. As the significant factors in determining the nature of the reflection are κ and $\frac{\sigma}{f}$ (σ in E.S. units) it follows that the relative effects of κ and σ vary greatly over the range of frequencies used in wireless.

The reflection conditions are, in general, quite different, depending upon whether the electric vector is perpendicular to the plane of incidence (Fig. 32a) or in that plane (Fig. 32b). The former case is simpler and will be dealt with first. If the surface were a perfect conductor then, for all values of θ, $(A\,|\underline{\phi}\,)$ would be -1; that is, the image dipole must be considered as carrying a current, equal in magnitude but opposite in phase, to that in the actual dipole. This is seen to be necessary to satisfy the boundary conditions because there cannot be any resultant electric field along the surface of a perfect conductor.

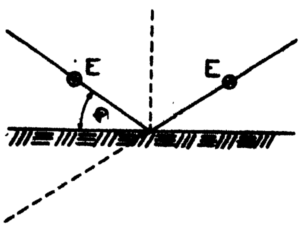


FIGURE 32a.

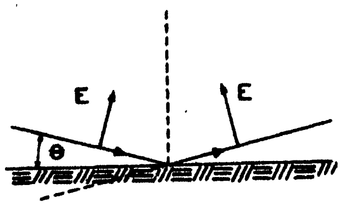


FIGURE 32b.

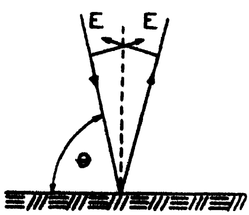


FIGURE 32c.

If the earth may be regarded as a perfect dielectric ($\sigma = 0$) then there will be a refracted ray passing into the earth, as well as a reflected ray. For small values of θ, $(A\,|\underline{\phi}\,)$ is very nearly -1. For larger values, ϕ is still 180°, but A depends upon the value of κ. Evidently, if κ was 1 there would be no reflection and A would be zero.

If we consider reflection at the surface of sea water ($\kappa = 81$, $\sigma = 10^{-10}$ E.S.U.), then for a frequency of 100 Mc/s $\left(\frac{\sigma}{f} = 100\right)$ the assumption that $(A\,|\underline{\phi}\,)$ is -1 is a very close one for all values of θ and for lower frequencies would be even closer.

For reflection at the surface of earth, assuming $\kappa = 10$, $\sigma = 10^8$ E.S.U., there is a much greater divergence however, and Fig. 33 (derived from McPetrie's curves[8]) gives values of A and ϕ for $\frac{\sigma}{f} = 1$, and $\frac{\sigma}{f} = 2{\cdot}5$, which correspond (with the values assumed for σ) to 100 Mc/s and 40 Mc/s respectively.

Consider now the conditions for reflection when the electric vector is in the plane of incidence (Fig. 32b and 32c). If the earth were a perfect conductor, then the horizontal components

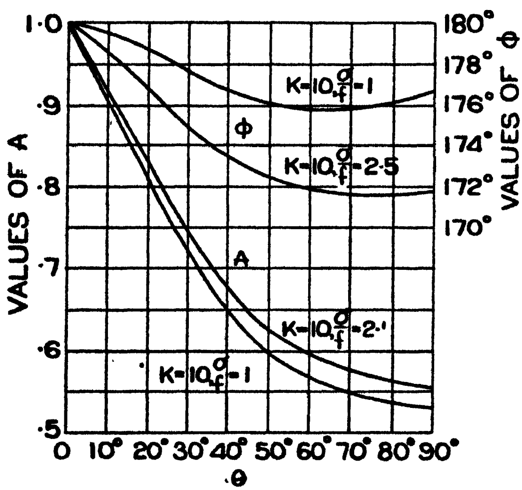


FIGURE 33.—Polarisation Perpendicular to the Plane of Incidence.

of the electric vectors would have to cancel for all values of θ. This requires that $(A\,|\underline{\phi}\,)$ would equal $+1$. It is evident that when θ is 90° there is no difference between the two polarisations. An examination of Fig. 32c shows that, with the conventional directions indicated, when θ is 90° and ϕ is 0°, the currents in dipole and image will actually be in opposite directions, thus conforming with the statement previously made for polarisation perpendicular to the plane of incidence.

If, on the other hand, the earth may be considered as a pure dielectric, then $(A\,|\underline{\phi}\,) = -1$ for small values of θ (as for polarisation perpendicular to the plane of incidence). As θ increases, however, A at first decreases, becoming zero (that

is, there is no reflected ray) at an angle called the Brewster angle, which depends upon κ, then increases again; ϕ is 180° below the Brewster angle and 0° above.

When actual values of κ and σ are considered, we obtain curves such as shown in Figs. 34 and 35, it being observed that with ultra-short waves we approach the condition of $\phi = 0$. From these curves it will be seen that the greater $\frac{\sigma}{f}$ is, relative to κ, the lower is the angle at which A becomes a minimum and ϕ changes quickly, but in all cases the image and actual

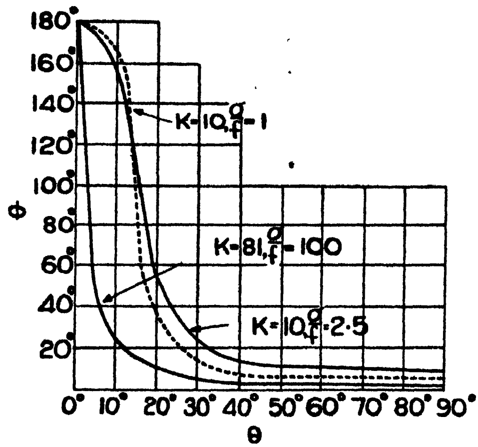


FIGURE 34.—Polarisation in the Plane of Incidence.

dipoles carry equal currents in opposite phase when $\theta = 0°$. It would therefore appear that wave propagation directly along the ground when σ is finite, is not possible, a result which we know to be contrary to experience. This discrepancy will be dealt with later.

The Ray Theory. We can now apply our discussion of reflection at the earth's surface to an approximate theory of propagation which is useful in certain cases.

The state of affairs existing when transmitting over a plane surface from an elevated transmitter to an elevated receiver may be seen by examination of Fig. 36. The attenuation of the direct ray will be entirely "geometrical," that is, due to its spread, and if T is a dipole then the amplitude of the direct ray will be inversely proportional to the distance.

If T is a horizontal dipole perpendicular to the plane containing the propagation path, then, whatever the value of θ, the incident wave is polarised perpendicular to the plane of incidence. By using curves in Fig. 33 we can therefore find the resultant field at R and we could plot a polar diagram for the horizontal dipole at T.

Consideration will show that, except for communication with nearby aircraft, or between very high, closely-adjacent hills, θ will not exceed about 10°. It follows from Fig. 33 that we may take $(A \underline{|\phi})$ as equal to -1 in many cases.

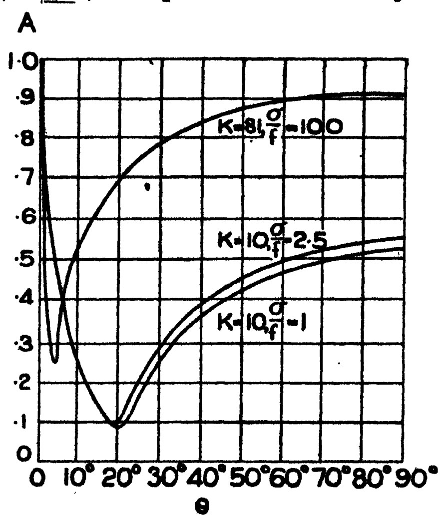


FIGURE 35.—Polarisation in the Plane of Incidence.

If T is a horizontal dipole but its axis is not perpendicular to the plane containing the propagation path, then the incident wave will contain components polarised in the plane of incidence as well as perpendicular thereto and it will be necessary to treat the components separately. The horizontal dipole has directive properties and if remote from earth does not radiate along its length. When near earth a vertically-polarised wave is radiated in this direction.

Suppose now that T is a vertical dipole. For all values of θ this will produce a wave entirely polarised in the plane of incidence, and by reference to curves such as those of Figs. 34 and 35 we can find the resultant field at R. For very small values of

θ we see that we can again assume that $(A \underline{|\phi})$ is -1, but it is evident that this approximation is only true over a very small range of θ, and, furthermore, may be invalidated by an effect to be discussed in the next section.

Since the values of κ and σ are so variable and are frequently not known at all closely, it is useful to have roughly approximate polar diagrams for a vertical dipole above the earth's surface. For values of $\frac{\sigma}{f}$ occurring in the short or ultra-short wave band, Fig. 34 suggests that for small values of θ the best simple assumption is to take $(A \underline{|\phi})$ as -1, whilst for higher angles $(A \underline{|\phi})$ should be taken as $+1$. Polar diagrams obtained in this way (but for a half-wave aerial, not a dipole) are shown in Fig. 92, page 178.

FIGURE 36.

Reverting now to a consideration of Fig. 36 we see that if θ is small and $(A \underline{|\phi})$ may be taken as -1, since the distance travelled by the direct and reflected wave is almost the same, the resultant field strength at the receiver will be much less than that due to the direct ray alone.

The field strength in volts per metre when the above conditions approximately obtain is:

$$E = \frac{240\,\pi^2\, I\, l}{\lambda^2\, r^2}\, h_T\, h_R$$

where l = effective length of T in meters.

I = current in T in amperes and all lengths are measured in metres.

Raising T or R (or both) will increase the resultant field and a reduction of wave length will do the same, since the difference in length of the two paths will become greater, as measured in wavelengths. Evidently, however, the formula can only apply if both aerials are raised somewhat $\left(\text{at least } \frac{\lambda}{2}\right)$ above earth.

The above formula must be used with caution and with a clear realisation of the assumptions upon which it rests. When valid this expression will be approximately true whether the dipole is vertical or horizontal, that is whether vertically or horizontally polarised waves are being used, but in the latter case it will only give the field strength in a direction normal to the dipole.

From the foregoing discussion it will be evident that the ultra-short waves will give good results if signalling is to be carried out between elevated sites such as between steep hills, because the effective values of h_t and h_r will be large.

A consideration of Fig. 37, however, will show that even in such a case, there may easily be more than one reflected wave. In the case shown, one reflected ray will largely neutralise the direct ray in the way previously discussed, but

FIGURE 37.

this leaves one reflected ray unopposed and therefore the field at R is much greater than indicated by the equation. It is, therefore, possible for the received strength to be large or very small and, when setting up a receiving site, the aerial should be moved about to find the best position.

It will also be evident that a small change in the transmission frequency will alter the length of the different paths, as measured in wavelengths, and will alter the phase-angles between the various components at the receiver and give a different resultant field-strength. If the transmitter is radiating a modulated wave, therefore, it does not follow that the various frequencies will preserve their correct relationship to the carrier, either in magnitude or phase. This effect is likely to be pronounced when very short waves are used for television.

The Surface Wave. All that has been discussed hitherto concerning reflection and propagation assumes that at the point where reflection occurs, the incident wave is a plane wave. If the dipole is sufficiently near to the earth (as

measured in wavelengths), however, the incident wave has not become plane before the surface is reached, and under these conditions the "image" theory will not give the total field.

In such cases there is a surface wave, the foot of which travels along the earth's surface. This wave undergoes attenuation because it produces currents in the earth, and there will be a wave travelling through the earth.

The attenuation depends in a complex way upon κ and $\frac{\sigma}{f}$ For the lower frequencies it is the conductivity that is important, the higher this is the lower being the attenuation. For frequencies within the short and ultra-short wave bands κ is also an important factor, the higher the value of κ the lower the attenuation. Thus for all frequencies the surface wave is less attenuated over sea than over land. The attenuation increases with frequency and becomes very great in the ultra-short wave band.

The magnitude of the surface wave from a vertical dipole is such that the ray theory previously mentioned ceases to be a useful approximation unless the dipole is about one wavelength above earth and it can be seen, therefore, that it is only likely to be of use for ultra-short waves.

When a horizontal dipole is used, however (and we are concerned with propagation in a direction perpendicular to the axis of the dipole), the surface wave is of much smaller magnitude and the optical theory with the image in antiphase is a useful approximation even when the dipole is only $\frac{\lambda}{4}$ above earth.

When a wave is travelling over an imperfectly conducting earth, the electric field must have a component along the direction of travel because the induced currents in the earth require potential differences over the surface to produce them. There will also be a wave travelling through the earth and this absorbs energy. The electric field of a vertically polarised wave is therefore no longer exactly vertical but is tilted, the foot lagging behind. The angle of tilt is a function of σ and κ and by measuring the tilt σ and κ can be deduced.

Another way of considering the matter is to realise that since there is a loss of energy at the foot of the wave there must be a

downward supply of energy from the upper part of the wave, to balance matters up.

The surface wave is able to travel beyond the optical range, that is, to follow the curvature of the earth, because of two effects: diffraction, and refraction in the lower air layers. It is well known that light rays "leak" around an opaque object and illuminate an area that should be in shadow if a strictly straight-line propagation only was considered. The magnitude of this diffraction effect increases with the wave length and it will be evident that diffraction will be an important feature of wireless wave propagation, since the waves are immensely longer than light waves.

Any theory allowing for diffraction must assume, of course, that the earth is a smooth sphere and hills or other obstructions may greatly influence the signal strength actually obtained. Such local obstacles will have more effect on the shorter wavelengths because diffraction is less marked, and the nearer the obstruction is to either transmitting or receiving aerial the more influence it will have.

The original diffraction theories were worked out for transmitting and receiving aerials on the earth's surface. Elevation to any appreciable fraction of a wavelength is clearly impracticable on the longer waves and would not be markedly beneficial.

As the wavelength is reduced, however, the range of a transmitter on the earth's surface to a receiver also on the earth's surface becomes exceedingly small and hence, when ultra-short waves are being used either transmitter or receiver (or both) will usually be raised to a height which may be several wavelengths.

It is evident that the "ray" theory cannot be employed for ranges greater than the optical because diffraction is not allowed for.

Eckersley's Theory of Ultra-Short Wave Propagation.[9] T. L. Eckersley has extended the diffraction theory to deal with the most general case of transmission from a raised vertical dipole, around the curvature of the earth (allowing for the earth's σ and κ) to a raised receiving dipole. The results of the theoretical analysis have been presented in a set of curves giving the field strength to be expected at different distances

from a transmitting dipole on the earth's surface, for different heights of receiving aerial. Alternatively, the curves will

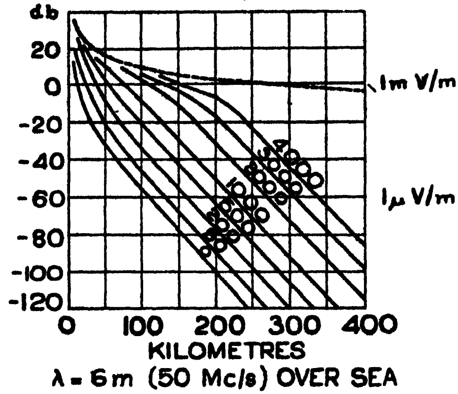


FIGURE 38.

give the field strengths at the earth's surface from a transmitting dipole raised to different heights. These curves, of which two sets are prepared, one for transmission over sea,

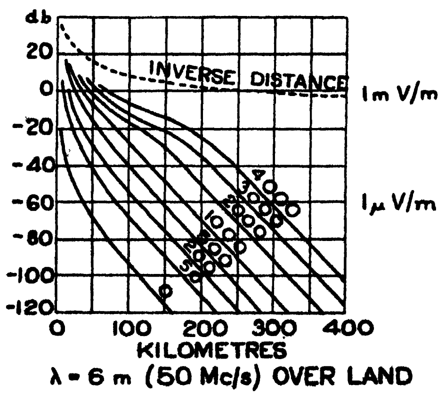


FIGURE 39.

and one for transmission over land, are shown on page 540, but Figs. 38 and 39 give one set for 6 metres, and it will be seen that the transmission of such waves over the sea (Fig. 38) is distinctly better than over land (Fig. 39), and that the field

strength obtained increases rapidly with height of transmitting (or receiving) aerial.

The use of the ray theory in cases where it is not even approximately true, has lead to the belief that signals will get suddenly weaker when we pass beyond the optical range. The full theory shows no such sudden change, except that at great heights (where the ray theory is more applicable) the curves do bend over and show increased attenuation after the optical range is exceeded.

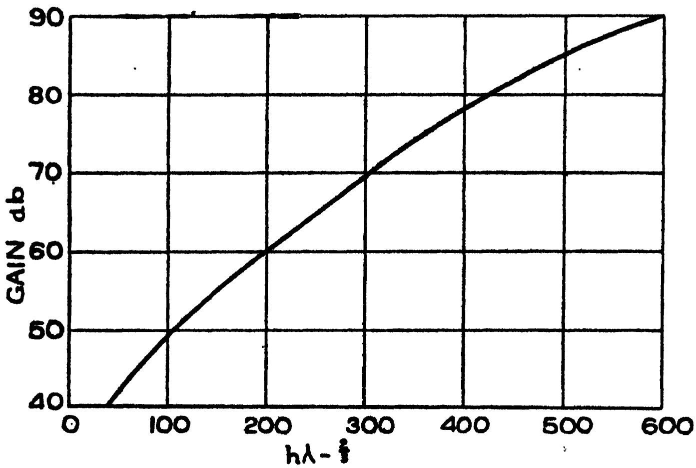


FIGURE 40.

Eckersley has produced an auxiliary curve giving the gain with height, so that a value of field strength can be obtained when both transmitter and receiver are raised. He finds that above a certain height, h_0, such that $h_0 \lambda^{-\frac{3}{2}} = 50$, the gain with height is practically independent of the earth's constants σ and κ and that it is therefore possible to give a curve of gain against $h\lambda^{-\frac{3}{2}}$ as Fig. 40. The gain below h_0 is, however, very dependent upon σ and κ and in order to find the total gain due to raising to a height h, it is necessary to subtract a constant number of dbs. from the given curve, in accordance with the table below :

Wavelength (Metres)		2	4	6	8	10
Dbs. to be subtracted	Over Land	0	2·2	3·7	4·9	5·9
	Over Sea	13·6	18·0	20·8	22·9	24·4

An example may make the use of the curves more clear :

> Required to calculate the field strength at a distance of 50 km. from a 50 watt transmitter employing an aerial (assumed to have a cosine polar diagram) at a height of 50 m. The receiving aerial is on an airplane at a height of 1,000 metres. The transmission is over land (assumed $\sigma = 10^{-13}$ e.m.u., $\kappa = 5$) and the wavelength is 6 metres.

From Fig. 39 the received energy on the ground at 50 km. from a 1 kW transmitter at 50 metres height is 44 dbs. below the datum (i.e. the energy corresponding to 1 mV/m.) The actual power being only 1/20th of that assumed in the curves will reduce the received energy by 10 log 20 = 13 dbs, thus making the F.S. —57 dbs.

For this case $h\lambda^{-\frac{1}{2}} = 303$ and from Fig. 40 the gain due to the height of the airplane is 71 dbs., from which 3·7 dbs. have to be subtracted (see Table), and hence the energy becomes

$$-57 + 71 - 3{\cdot}7 = 10{\cdot}3 \text{ db. above datum.}$$

If the field strength in mV/m is "x," then :

$$+10{\cdot}3 = 20 \log x \text{ from which}$$
$$x = 3{\cdot}3 \text{ mV/m.}$$

In Appendix I will be found a series of curves for wavelengths between 2 and 6 metres.

Refraction of Waves in the Lower Atmosphere. The refraction effect due to changes in the lower air layers also affects the wave. When an E.M. wave passes from one medium to another the following relationships exist between the incident and refracted waves. There will, in general, be a reflected wave in the first medium, which is not considered here.

$$\frac{\text{Sin } \phi_1}{\text{Sin } \phi_2} = \frac{v_1}{v_2} = \sqrt{\frac{\mu_2\kappa_2}{\mu_1\kappa_1}} = \sqrt{\frac{\kappa_2}{\kappa_1}} \text{ if both media have the same}$$

value of μ_1, which is usual. (ϕ_1 and ϕ_2 are the angles between the normal and the incident and reflected rays, respectively.)

In the case of dry air, κ is proportional to the quotient of pressure and absolute temperature. Both of these normally decrease with height in the case of the atmosphere but in such a way that their quotient decreases with height, and hence κ does the same.

Suppose a wave to be travelling in a direction tangential to the earth's surface, then the lower part of such a wave will be travelling at a lower velocity than the upper part and the wave will be bent towards the earth and refraction will therefore tend to increase the range obtainable.

In the early days of long-wave wireless it was thought refraction might account for the ranges obtained, but such an explanation proved quite inadequate and we can say that refraction plays a negligible part in long wave propagation.

Where ultra-short waves are concerned, however, because diffraction is much less, refraction becomes important in increasing the range, although to what extent is not very clearly established as yet.

Because the dielectric constant of water is so large (approximately 80), the actual value of κ of the lower atmosphere is very dependent upon the amount of water vapour present. This is, of course, a very variable quantity but is definitely greater in summer than in winter. The amount of water vapour present decreases rapidly with height in general, and, therefore, this still further increases κ for the lower atmosphere compared with the value for the higher and the waves are bent to an increased extent.

The path of a wave through an atmosphere having likely average properties has been worked out and found to have a radius of curvature about four times that of the earth (23,800 km. in summer and 26,500 km. in winter) and it has been shown that the effect of refraction may be allowed for theoretically by replacing the true radius of the earth (6,370 km.) in the diffraction formulæ by an increased radius (8,650 km. in summer, 8,420 km. in winter). Such calculations can, of course, give only an estimate of the effect of refraction, since the composition of the atmosphere is so variable. It occasionally happens that the temperature of the air at the earth's surface is lower than at some distance above the surface. When this occurs the effect of refraction is usually to considerably increase the signal strength at the receiver.

Comparatively sudden local variations in the atmosphere result in ultra-short waves usually suffering from slow, deed fading at the longer ranges where refraction is playing an important part in bringing the energy down to the receiver.

It will be evident that comparatively small "pockets" of air having a different dielectric constant from the surrounding air (due, for example, to their being warmer) may change the apparent direction of an ultra-short wave considerably, whilst they would have no effect upon a longer wave.

The theoretical work on short and ultra-short wave transmission (excluding transmission through the ionosphere) may be summarised as follows:

The range of a transmitter on the earth's surface to a receiver also on the ground, becomes smaller as the wavelength is reduced. Under certain conditions theoretical curves are available from which the field strength at an elevated receiver, produced from an elevated transmitter, may be deduced. The earth's surface has to be assumed to be smooth and spherical and likely average values of κ and σ assumed. Refraction may produce an appreciable increase in signal strength at longer distances and can be expected to produce fading effects similar to those experienced with longer waves from the ionosphere. An approximate "ray theory," though of limited application, is useful and simple in certain cases and shows us that where communication takes place over uneven ground or obstructions commensurate with the wavelength, there are frequently several paths by which energy can reach the receiver and in consequence an interference pattern may be produced.

We now propose to study very briefly some of the more important experimental investigations which have been made in order to see how far the theories outlined are correct and complete.

Experimental Studies of Transmissions of 5 to 7 metre Waves.—Very many investigations have been made on these wavelengths either to check the theories discussed or to provide data for the establishment of a commercial service.

All investigators have found that signal strength varies rapidly with receiver position as the "ray" theory suggests. In endeavouring to check theoretical field strength/distance curves, therefore, it is necessary to choose as open a site as possible and it may be necessary to take several readings at sites very near together and take a mean. Any site chosen must be well away from all power or telegraph lines, railways

and trees, and should preferably be uniform in character and flat.

Many measurements [10] have been made on transmissions at several frequencies in the neighbourhood of 50 Mc/s (6 metres) from the Empire State Building, New York, the aerials of which are the highest fixed aerials in the world (400 metres above earth) and the results obtained and published by Jones are in good agreement with the Eckersley diffraction curves.

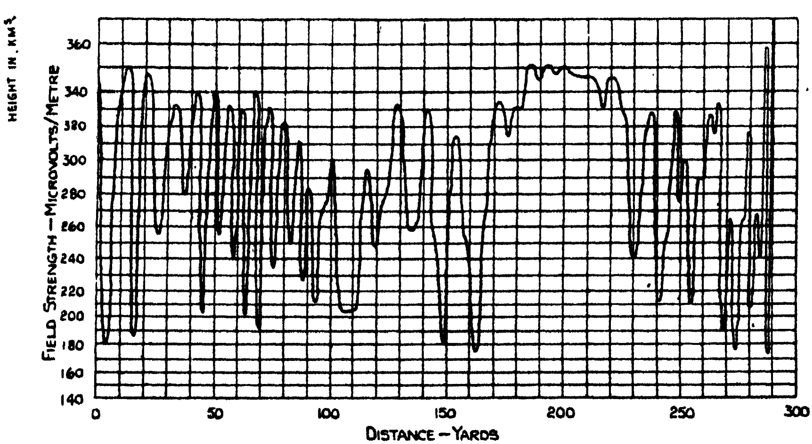
Jones [10] and other workers have observed the large fluctuation in signal strength which can be produced by moving cars, etc. At these wavelengths it is evident that such objects are sufficiently large in comparison with the wavelength to re-radiate to a considerable extent. It has also been observed that aircraft can produce noticeable fluctuations of signal strength, even when at some distance from the receiver. Since the aircraft is in a much stronger field than the receiver (due to its height) re-radiation from it is considerable. As an instance, television reception at Chelmsford (30 miles from the transmitter) often suffers rapid fading due to aircraft between Chelmsford and London, the picture brightness varying as the changing position of the aircraft changes the phase of the re-radiated wave relative to the direct wave. Tests in rural districts also show rapid variations of signal strength with position, especially near trees. A typical curve obtained by the B.B.C. (Fig. 41) shows the large variations obtained for small changes of position.

Maclean [11] studied 50 Mc/s transmissions at three sites, one of which was within the "optical" range, one 700 feet below the line of sight, and one 11,400 feet below. He found that even within the optical range there was slow fading up to 10 dbs. in amplitude, whilst at the longer distances fading was much more pronounced, the maximum values of the field strength being about 20 dbs. above the average value. There was no correlation between the fading at the three sites. The average field strength was much higher at night, probably due to refraction.

It has been demonstrated by George [13] that the frequency response curve of a wide-band receiver (such as for television) can be greatly modified by multiple-reflection effects.

An extensive study [16] of the transmissions on 6·67 metres (45 Mc/s) (vision) and 7·25 metres (41·5 Mc/s) (sound) from the B.B.C. Television Station at Alexandra Palace, London, has been made and the B.B.C. have published "contour" maps, one of which is shown in Fig. 42.* From these maps, curves for field strength in the different directions, N. S. E. and W., have been plotted in Fig. 43, and compared with Eckersley's theoretical curve, showing the close agreement.

The British Post Office [17] have made a number of investigations and report that over a "non-optical" path of 85 miles, fading may be as high as 60 dbs. at 60 Mc/s.



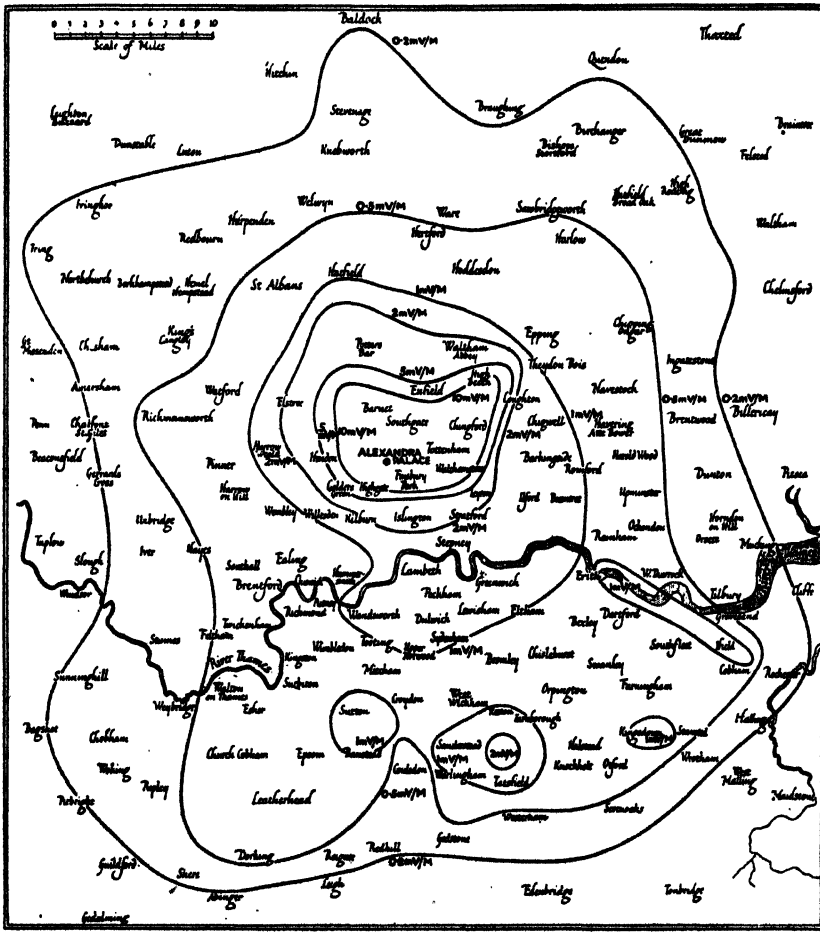
* FIGURE 41.

The gain of signal strength with height above the earth has been measured by a number of observers, and in particular the work of Jones [10] using an auto-gyro aircraft have been shown to agree closely with Eckersley's theoretical curves.

Experimental Studies of 1 to 3 Metre Waves. The ray theory in the form given by McPetrie and Saxton [6] has been checked by them, using wavelengths of 2 and 3 metres and very good agreement has been obtained. Conditions were such that the ray theory, with image in anti-phase, was a reasonable approximation. It was confirmed that on an

* By permission of the British Broadcasting Corporation.

open site, vertically and horizontally polarised waves gave the same field strength and that signal strength was directly proportional to receiver height as stated in the theory. It was



* FIGURE 42.

observed that the 2-metre transmissions were attenuated to a greater extent when passing over London, but the 3-metre transmission appeared to be unaffected.

Continuously during one year, Burrows[12] and his associates studied the propagation of a 2-metre (150 Mc/s) wave over a

* By permission of the British Broadcasting Corporation.

"non-optical" 60 km. path and found fading up to 20 dbs. The average field strength was higher during the night but the fading more pronounced. Conditions were very similar at a distance of 200 km.

Experimental Studies of Waves Shorter than One Metre. Some of the pioneer work on these wavelengths was done by Yagi and Uda [22] in 1928, and in 1931 the I.T. and T. Corp. [23] commenced to study the propagation of 18 cm.

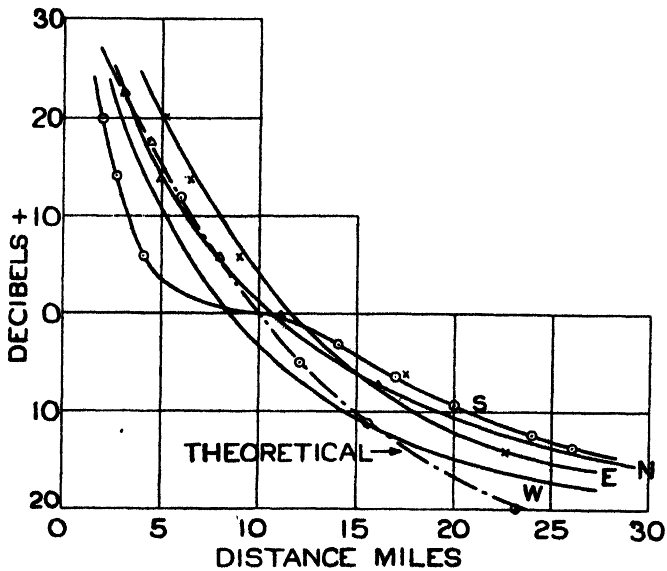


FIGURE 43.

waves across the Straits of Dover. This is probably the shortest wavelength for which there is a published account of propagation conditions. In these tests between fixed stations well within the optical path signals were steady when weather conditions were steady and were unaffected by fog, rain, or snow. Abrupt changes in temperature and pressure often produced fading up to 40 dbs., particularly on still summer days. The results suggested that the direct ray was fading because under these weather conditions the atmosphere was very inhomogeneous and changes in refractive index caused the ray to be deflected. In addition, there was almost certainly a reflected ray from the surface of the water and due to changes

in the atmosphere through which the rays passed, the phase difference between direct and reflected ray was continually changing. There was some evidence that the rise and fall of the tide affected the signal by altering the length of the reflected ray path.

A 6-metre circuit which was established between the same two points for comparison showed negligible fading so that it was clearly established that the very short waves were much more affected by changes in the lower atmosphere. It was noticeable that reception on 18 cm. was unaffected by ignition noise, so noticeable on the 5 to 7 metre band, or by thunderstorms.

Marconi and his assistants[25] conducted an extensive investigation of 50 cm. waves in 1932, using the yacht *Elettra* as a receiving station and employing sharply directional transmitting and receiving arrays. On one typical test over sea, good telephone signals were received at 58 miles (the optical range being about 52 miles) but very deep fading. allowing only of occasional reception, was then experienced up to 87 miles. At this distance signals improved considerably and remained good up to 100 miles, being finally lost at 110 miles. During a later test, consistent reception was possible at distances up to five times the optical range even when hills intervened, and signals were reported at a distance of nine times the optical range.

Hersberger[24] made a brief test using a 75 cm. transmission and obtained useful telegraph signals at 88 miles over sea, this being five times the optical range. Up to 20 miles the signals were strong and steady, from 20 to 30 miles became weak and fading, and above 30 miles signal strength varied little with distance and there was some fading, these results agreeing with Marconi's. The long ranges so obtained by these experimenters have not been satisfactorily explained by existing theories.

Application of Short and Ultra-Short Waves over Short Distances. The available band of short wavelengths is so valuable for long-distance communication that the use of such waves for short distances must be limited. In tropical countries, however, long and medium waves are so seriously interfered with by atmospherics that short waves

are sometimes employed for broadcasting or comparatively short distance communication.

Short waves are also employed for such services as police wireless, where only a small area has to be covered. In this case, the main advantage compared with longer waves is that the very small aerial which can be carried by a car is a more efficient radiator of the short waves.

This advantage is, obviously, still greater if ultra-short waves are used, and interference is not created at great distances as it is with short waves. The principal difficulty of the ultra-short waves is the great variability of the signal with change of receiver position, but in spite of this ultra-short waves are finding increasing use for such services.

Ultra-short waves are very useful for point-to-point communication over short distances, especially if the transmitter and receiver can be on an elevated site. The radiation can readily be confined into a narrow beam, interference is small and the equipment compact. The British Post Office have a number of such circuits in use, linking the outlying islands to the telephone network at a very small cost compared with a cable. A number of circuits, additional to those carried by cable, are also provided to Northern Ireland. Some of the P.O. circuits are over distances considerably greater than the optical range.

Waves below one metre have hardly been used for commercial services. Only two such circuits are known to have been installed, one on 50 cm and the other on 18 cm.

Selected references are given at the end of Chapter V.

CHAPTER V

THE PROPAGATION OF SHORT AND ULTRA-SHORT WAVES THROUGH THE IONOSPHERE

When, in 1901, Marconi successfully demonstrated that communication across the Atlantic (actually between Poldhu, Cornwall, and Newfoundland) could be achieved with electromagnetic waves, not only did he place the first milestone on the road to long distance wireless communication, but he opened up a new field of thought regarding our earth's surrounding atmosphere and stimulated scientific interest in a direction which has continued up to the present day.

For this experiment a wavelength of about 1,300 metres was used and the theories outlined above are insufficient to explain the long range obtained. In fact the theoretical figure for the field strength at a great distance on a perfectly conducting spherical earth was only 10^{-8} of that probably obtained.

Assuming the earth to be a spherical conductor, Heaviside (and Kennelly) conceived an upper ionised layer forming a second spherical conductor concentric with earth, between which is a homogeneous insulating medium. These two spaced conductors thus form a spherical transmission line, so that any electrical disturbance across the line creates plane polarised electro-magnetic waves which are propagated through the insulating medium between these boundaries, and such waves, therefore, automatically follow the earth's curvature.

For a number of years there was no very direct experimental evidence for the existence of this ionised layer except that crude measurements of field strength showed that this was of the order predicted by approximate theory. From 1925 onwards, however, many workers have studied the problem by various methods. As a result, it has been found that

there are, in fact, several effective "layers" of ionised air and the whole region has been termed the ionosphere. In this book only a simple treatment of the subject will be attempted dealing particularly with practical application to wireless communication, and the reader is referred to the very extensive literature (references to which are given at the end of this Chapter) for more detailed information.

Before proceeding to discuss the ionosphere and the behaviour of an electro-magnetic wave passing through it we may do well to revive our knowledge on two points.

Phase and Group Velocity. The ionosphere is a type of medium in which the velocity of waves having different

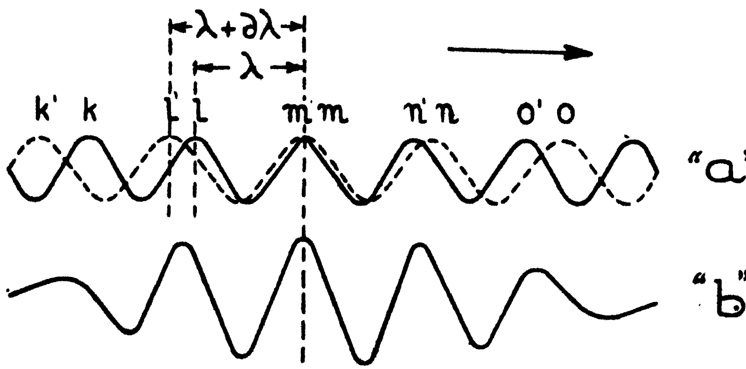


FIGURE 44.

frequencies is different. In such a medium it is necessary to distinguish between phase and group velocity. It has already been stressed that all communication requires more than one frequency, the only wave comprising only a single frequency being a sinusoidal wave continuing indefinitely. If we have actually a short wave-train or group (such as forms the "dot" in C.W. telegraphy) then this may be considered as due to a number of infinite wave-trains. Consider the group of Fig. 44b, made up of the infinite wave trains of Fig. 44a, having wavelengths of λ and $\lambda + d\lambda$. In a simple medium any wave crest in either of these waves will be travelling with a velocity v, and hence the same wave crests always continue to form the peak of the group, and the group is also moving at v. In such a medium, therefore, we do not need to distinguish between

the velocity of a single wave crest (the phase velocity) and the group velocity.

If now our group enters a medium where phase velocity varies with frequency (an ionised medium being one example of this), the velocity of the wave-train depicted by the full curve Fig. 44a may be v, but that of the other may be $v+dv$. Consequently the crest l' will in time catch up on the crest l, and hence the maximum value of the group now corresponds to the position of ll'. It will be seen that the group has fallen back with respect to the individual waves; in other words, the group velocity is less than the phase velocity.

The relative velocity of l' with respect to l is dv, and hence it will take a time $\frac{d\lambda}{dv}$ to catch up on l. When this has occurred the group will have moved back λ, relative to individual wave crests. Hence the velocity of the group is given by

$$v_g = v - \lambda \frac{dv}{d\lambda}$$

For the case of a group of waves passing through an ionised medium, therefore, the phase velocity will be greater than c, the velocity of light in a pure dielectric, because the refractive index is less, but the group velocity will be less than c, and hence if we measure the time taken by a group to pass through the ionised medium it will be found to be longer than if it had travelled with the velocity c. The reduction in the group velocity is dependent upon the electron density of the medium through which the group is travelling.

In the case of transmission through the ionosphere, the phase velocity will be higher than that of light whilst the group velocity will be lower.

Polarisation of E.M. Waves. We have already noticed that a vertical dipole set up on the earth's surface produces a vertically polarised wave, that is, one in which the electric field is vertical. The magnetic field is horizontal, so that the direction of propagation, the electrical field and the magnetic field, are mutually perpendicular.

After the passage of a wave through the ionosphere it may be found that the plane of polarisation is revolving. If the value of the electric field in this plane remains constant, it

may be resolved into two equal components at right angles to each other and in time quadrature. The magnetic field consists of two similar components and the wave is said to be *circularly polarised.*

If the electric field is represented by a vector, say, OP, Fig. 45, then if the wave is circularly polarised, this vector must be regarded as rotating at a uniform rate as it progresses forward. Thus its extremity traces out a spiral, and the projection of this spiral on a plane normal to the direction of propagation will be a circle as shown. Its components along OY and OX are therefore given by $OP \sin \omega t$ and $\pm OP \cos \omega t$, respectively, the sign of the second component depending upon the direction of rotation of OP.

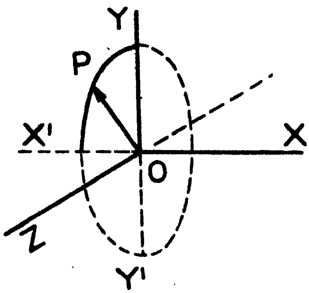


FIGURE 45.

The amplitude of OP may vary as it revolves, so that its extremity traces out a spiral whose projection is not a circle but an ellipse, and the wave is then said to be elliptically polarised, circular polarisation being merely a particular case of this. As before, this resultant can be resolved into two components along OY and OX, but these two components will no longer be of equal amplitudes.

The Passage of an E.M. Wave through the Ionosphere. Suppose an E.M. wave having an electric field $F \sin \omega t$ to be traversing ionised air in which there are N free electrons per c.c. Let e be the charge on an electron and m its mass. The electric field will exert a force $Fe \sin \omega t$ on the electron and, if v is the velocity produced, then

$$m\frac{dv}{dt} = Fe \sin \omega t \text{ and } v = -\frac{Fe}{\omega m} \cos \omega t.$$

The motion of N electrons per c.c. constitutes a current density $N e v$ or

$$-Ne\frac{Fe}{\omega m} \cos \omega t$$

In a pure dielectric the displacement current produced by $F \sin \omega t$ would be

$$\frac{\kappa}{4\pi} F\omega \cos \omega t$$

and the total current density is, therefore,

$$\frac{\omega F \cos \omega t}{4\pi}\left[\kappa - \frac{4\pi N e^2}{\omega^2 m}\right]$$

If κ is taken as unity for unionised air, then the dielectric constant of the ionised air is

$$1 - \frac{4\pi N e^2}{\omega^2 m} \quad \text{or} \quad 1 - \frac{N e^2}{\pi f^2 m}$$

The refractive index of the ionised air (relative to unionised air) will be given by

$$\mu = \sqrt{1 - \frac{N e^2}{\pi f^2 m}}$$

If we substitute values for e, m and π

then $$\mu = \sqrt{1 - 8{\cdot}1 \times 10^7 \frac{N}{f^2}}$$

A wave entering the ionised air will therefore be refracted away from the normal to the surface as shown in Fig. 46. If N is increasing with height, then the wave will travel in a curve, and if the bending is sufficient, the wave will be directed back to the earth again.

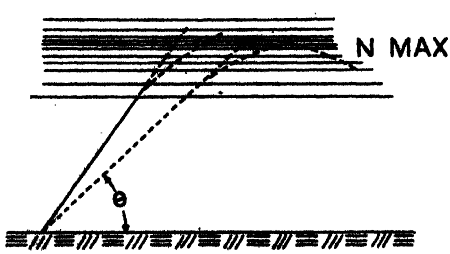


FIGURE 46.

We cannot trace out accurately the path of a wave through the ionosphere because we do not know sufficiently well how N varies with height.

The velocity of an E.M. wave is given by $\frac{1}{\sqrt{\mu\kappa}}$ (where μ here is permeability) and if c is the velocity in un-ionised air, the velocity in the ionised air becomes $\frac{c}{\sqrt{1 - 8{\cdot}10 \times 10^7 \frac{N}{f^2}}}$ and the group velocity becomes correspondingly lower than c. It is to be noted that the velocity now depends upon f.

As we make our wave enter the ionosphere more steeply it evidently requires to be bent through a larger angle if it is to be returned again. No rays will be returned above a certain angle given by $\cos\theta_o = \sqrt{1 - 8{\cdot}10 \times 10^7 \frac{N_{max}}{f^2}}$ though, if the frequency is low enough, θ_o may be 90°. As the frequency is raised we eventually reach a value for which no wave is returned from the ionosphere, however oblique the incidence.

A consideration of the geometry will show that, due to the earth's curvature, even a ray which leaves the transmitter tangentially to the earth's surface cannot enter the ionosphere at less than a certain value of θ, which depends upon the height of the lower edge of the ionosphere. It will be seen, also, that if a ray travels in a straight line it will make a continually increasing angle with the tangent to the earth below. The smallest angle which a ray can make with the E layer is about 8° and with the F layer about 14° (see page 92).

To gain some idea of numerical values we may note that if N is 4×10^5 electrons per c.c., and f is 20 Mc/s, then $\mu = 0{\cdot}96$, phase velocity 1·04 and group velocity 0·96.

If N_{max} is 4×10^5, then the critical frequency for vertical incidence will be 5·7 Mc/s, whilst the highest frequency which would be returned for $\theta = 10°$ would be 33 Mc/s (9·1 m.). These values are approximate, due to the simplification of theory.

In the above discussion we have ignored certain factors and it now becomes necessary to discuss these. It will be evident that positive ions in an ionised gas will also experience a force on the passage of a wave. Their mass being so very much greater than that of the electrons, however, their motion is small and their contribution to the total current negligible.

The Effect of Collisions in an Ionised Region. When the electrons are set in motion by the wave, collisions will be caused with the gas molecules and the motion of the electrons will be modified. For the highest frequency the average time between collisions is so long compared with the period of the wave that the effect of collisions is not so very important. In our simple analysis the electron motion was such that an entirely quadrature component of current was added to the displacement current and hence κ remained a simple quantity and there was no absorption of energy in the ionosphere, any more than in un-ionised air. When the effect of electron collisions is allowed for, however, κ becomes a complex quantity. The total current in the ionosphere has now a component in phase with the electric field, namely a conduction current, and this draws energy from the wave, which is therefore attenuated. The calculations to obtain the effective dielectric constant are now much more complicated, but we can gain an idea of the result by considering the average effect of many collisions as providing a damping force proportional to the motion of the electron but opposing its motion. We have something akin to the mechanical vibration of a body having inertia and friction under the influence of a simple harmonic force. When there are a number of cycles of the wave during the average period between collisions, the frictional force is small and the electron motion is nearly in quadrature with the applied electric force as we have already seen, that is to say the ionosphere is acting as a pure dielectric.

If, however, the frequency is made much lower, the frictional force is large and the electron motion becomes nearly in phase with the electric field, that is the conduction current predominates and at such frequencies the ionosphere behaves as a good conductor rather than a dielectric. In this case an incident wave is therefore reflected and only a small amount of energy, which is quickly dissipated, enters the layer.

Thus the average collision time of the layer marks a dividing line between two types of wave propagation. This time is such that the medium broadcast frequencies lie near the dividing line and thus will be expected to have characteristics common to both long and short waves.

The Effect of the Earth's Magnetic Field. A second fact that has been ignored is that the ionosphere is situated in the earth's magnetic field. In consequence, if an electron is set in motion by the electric field of a wave, it may have another force acting on it due to the earth's magnetic field. If we consider a plane polarised wave entering normally a uniformly ionised medium, two special cases arise according as the direction of propagation is along or transverse to the direction of the magnetic field.

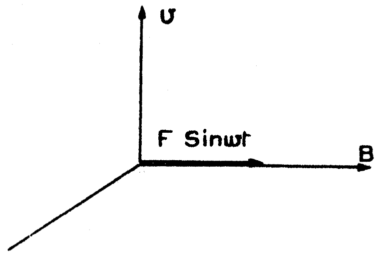


FIGURE 47a.

In the former case, shown in Fig. 47b, the electron motions become elliptical instead of "to and fro" as previously assumed, and it can be shown that, as a result, two circularly polarised waves are produced which travel with different velocities through the ionosphere and suffer different attenuations. In the case when the direction of propagation is transverse to the earth's magnetic field, two particular cases are shown in Figs. 47a and 47c.

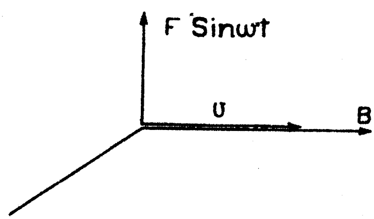


FIGURE 47b.

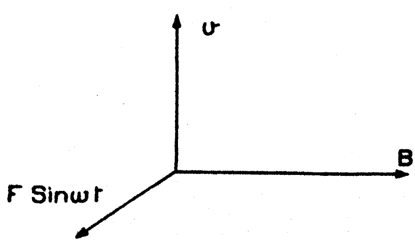


FIGURE 47c.

In 47a is the electric vector in the incident wave is along the direction of the field. The motion of the electron is therefore along the magnetic field and is therefore unaffected by it, and as though the earth's earth were absent.

In Fig. 47c the electric vector is transverse to the earth's field, and the motion of the electrons becomes elliptical in the plane at right angles to the field. The wave remains plane

polarised, but with a velocity modified by the presence of the earth's field. If the electric vector is neither transverse nor along the earth's field the wave is resolved into two plane polarised waves, one of the type in 47a and the other of the type in 47c.

In the general case when the direction of propagation is oblique to the direction of the earth's field the wave is resolved in the medium into two elliptically polarised waves (whose characteristics depend also upon the density of medium), and in the case of wireless transmission through an ionosphere of varying electron density, in which the direction of the wave is altering with respect to the direction of the field as the wave proceeds, the polarisation characteristic will change from point to point along the ray path. The polarisation of the wave on emergence will therefore depend upon the angle the ray makes with the earth's field as it leaves the ionosphere.

As the frequency of the wave is varied the elliptic paths of the electrons vary in shape because, for the same value of F the forces on the electrons due to F and B change in magnitude. At a certain frequency termed the resonance frequency, the elliptic orbits become very large and hence the number of collisions is very great, and the losses are greater than at other frequencies. For a value for B of 0·5 the resonance frequency is 1·40 Mc/s (214 metres).

At lower frequencies the magnetic field reduces the amplitude of the electron motions, thus reducing the collisions and hence the attenuation. At higher frequencies the main effects are those already discussed and at the shorter waves, especially for the very oblique incidence such as obtains with long-distance transmission, the earth's field has little effect.

Measuring the Properties of the Ionosphere. Existing methods of investigating the ionosphere are only capable of measuring the heights of the maximum electron densities. The properties have only been investigated consistently at a few places on the earth, the most intensive work having been done by the Bureau of Standards, Washington, U.S.A., the Radio Research Board at various places in England and by the Marconi Company, Chelmsford, England. Studies have also been made in India, Australia, the Polar Regions, and at Huancyo, Peru. The last named place was chosen because

it is on the earth's magnetic equator, and this simplifies in some ways the interpretation of the results obtained.

The existence of the ionosphere was indirectly shown by the fact that the various series of long wave signal strength measurements agreed reasonably well with theories dependent upon a spherical conducting "ceiling" around the earth, but the first direct measurements from which the effective height could be estimated were made by Appleton and Barnett in 1925 on about 300 metres. A transmitter was set up whose frequency was continually varied, and the received signal strength at a point 88 km. away was found to vary through a number of maxima and minima as the phase difference between the surface ray and the downcoming ray varied. When investigations of layer height were first commenced, the existence of an *F* layer was not suspected, but Appleton found a sudden discontinuity in measured height at certain times and deduced the existence of an upper layer.

Breit and Tuve in America first developed a method of measuring layer height which was adopted by Appleton and has since been used by most workers on the subject. A very short "pulse" signal is transmitted and is received at a station a mile or two away. The surface wave is therefore received and also a ray which leaves the transmitter almost vertically, a small portion of which may be "reflected" from the ionosphere. By the use of an oscillograph at the receiver, the time elapsing between the arrival of the surface wave and the reflected ray can be accurately determined and their relative amplitudes compared.

It is necessary to understand clearly the meaning of the height of a layer deduced from such measurements. Whilst the reflected ray is in the layer its group velocity will be reduced below c, whilst the velocity of the surface wave will equal c throughout its journey. If an equivalent path be assumed for the reflected ray which it is supposed to traverse with velocity c, then if the length of this path be l and the surface distance s, we have the relation $\frac{l-s}{c} = t$ where t is the interval between the reception of the two signals. It can be shown that the equivalent path is the "optical path" *TAPCR* (Fig. 48), so that the equivalent height measured is *MP*.

As the frequency on which the pulses are transmitted is made greater, we shall reach the critical frequency for vertical incidence. When θ_0 is 90°, the expression on p. 85 for $\cos\theta_0$ shows that $N_{max} = 1{\cdot}24 \times 10^4 \times f_c^2$, where f_c is the critical frequency in Mc/s.

Thus exploring the ionosphere with a variable frequency, it is possible to obtain the critical frequencies of the various layers as they become transparent at any time and place to waves of vertical incidence, and Fig. 49 shows a typical critical-frequency height curve taken at Chelmsford, summer, midday.

It will be noticed from Fig. 49 that the critical-frequency height curves split into two branches as the critical frequency is

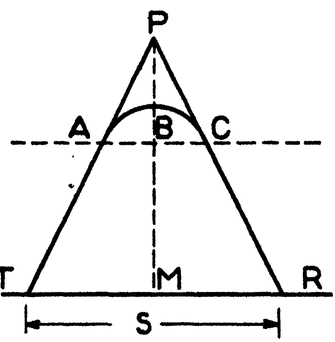


FIGURE 48.

approached. This is due to the "echo" being split into two signals separated from each other by the action of the earth's magnetic field. They are known as the ordinary and extraordinary rays, and the separation of the critical frequencies is dependent upon the strength of the earth's magnetic field. In the northern hemisphere, at vertical incidence, the ordinary ray emerges from the layer with a left-hand sense of polarisation when observed from above and the extraordinary ray with a right-hand sense, whereas in the southern hemisphere the reverse is the case.

It is evident that in wireless communication we are usually interested in the behaviour of waves incident obliquely on the ionosphere, whereas the pulse measurements deal with vertical incidence. The pulse measurements have also to be made at

lower frequencies than many of those used for long distance communication.

A very approximate relation between the two is seen from page 85, but methods for more accurate prediction of the behaviour of obliquely-incident rays from the data given by pulse measurements have been developed by Martyn, N. Smith, and Millington.

The first observations of oblique rays were those obtained by T. L. Eckersley, using the facsimile-telegraph apparatus. By its use he was able to measure the time interval between

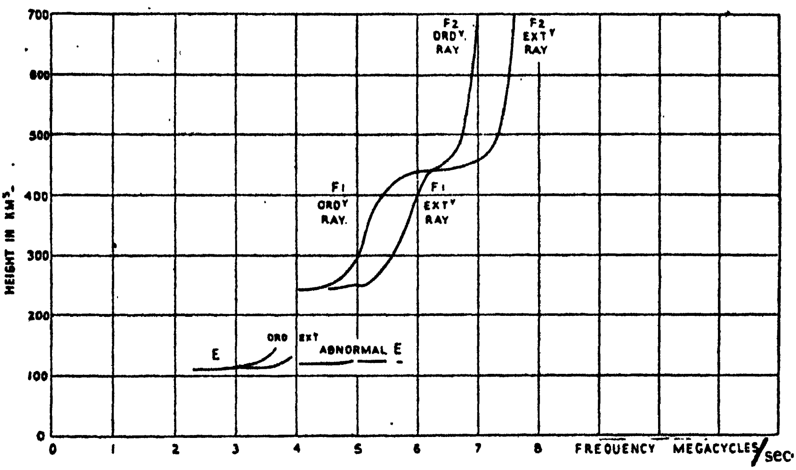


FIGURE 49.

different rays arriving at a receiver after having been reflected a different number of times from the earth (see page 100). From these measurements the maximum electron density in the layer concerned may be deduced.

The Structure of the Ionosphere. Having considered the methods by which the ionosphere has been "explored" we now turn to consider what these methods have revealed.

The ionisation of the atmosphere is brought about by ultra-violet light and corpuscular radiations from the sun. Hence, the number of free electrons present will increase in any part of the ionosphere which is exposed to sunlight. During hours of darkness, recombination will be taking place, that is, free electrons will be uniting with positive ions to form

neutral molecules, and at low atmospheric pressure this recombination process is slow and hence the density of free electrons will decrease continuously during the hours of darkness.

The density of free electrons at any time and place varies with height above the earth's surface and experimental evidence suggests the presence of several layers. Or more precisely, the number of free electrons per c.c. rises to more than one maxima with height as shown pictorially in Fig. 49, where frequency in Mc/s is proportional to electron density. These layers are enumerated E, F, etc., the notation being introduced by Appleton, who first discovered the F layer.

There are two principal layers in the ionosphere, namely E and F, the latter, however, often separating into two parts, the F_1 and F_2 layers respectively.

The lower, or E layer, remains constant in height, at about 100 km., but its density varies with the sun's altitude. Thus it will be most dense at midday in summer, and of minimum density during the night. Similarly the F_1 layer height remains constant at about 200 km., although it is not always observable as a separate layer owing to its merging with the F_2 layer at certain times; but, as far as can be judged, its density, like that of the E layer, can also be directly correlated with the sun's altitude. It is assumed that both the E and the F_1 layers are formed by ultra-violet radiation from the sun, as both these layers are absent during the winter periods at the Polar regions.

The F_2 layer does not appear to be caused entirely by ultra-violet radiation as it can be observed in the polar regions during the winter months, and its general behaviour is abnormal. First of all, its height is greatest in the summer daytime, and it descends to the level of the F_1 layer in winter. Secondly, the density of the F_2 layer in contrast with the E and F_1 layers is found to be lower in the summer months than during the winter and it also suffers from an annual as well as a bi-annual effect. Thus it is found that the height of the F_2 layer is decreased and its general density increased in *both* northern and southern hemispheres, between September and March. These inconsistencies of the F_2 layer are seen in Fig. 50, which correlates the equivalent layer density with years.

It is sometimes found that the density of the E layer increases

greatly for a short time. These sporadic conditions are local in character and the agent producing them is not definitely known.

Besides the principal layers mentioned, there is evidence of layers lower than the E layer, known as the C and D regions. The existence of a G layer beyond the F_2 layer has been suggested as an explanation of some echoes which show a very great equivalent height, but other more probable explanations

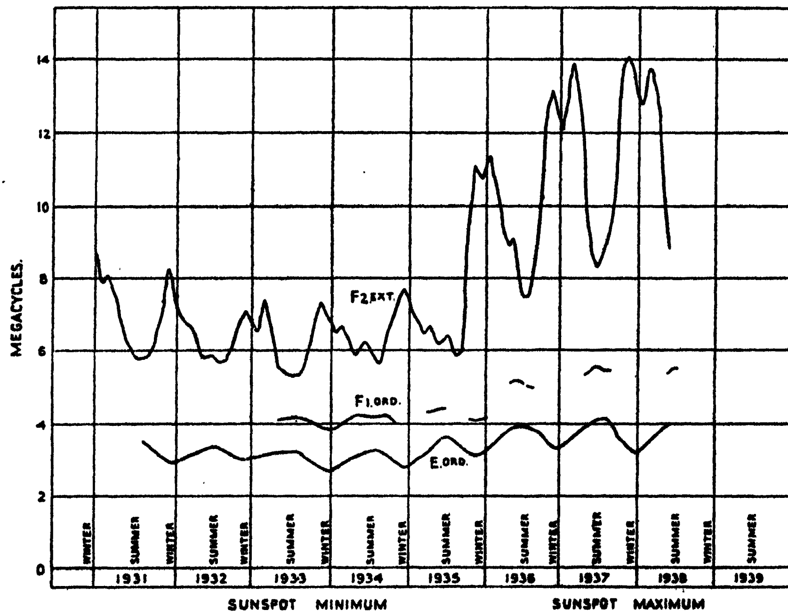


FIGURE 50.

have also been advanced. There is some evidence for a D region at between 50 and 90 km., and a C region between 20 and 35 km., both these layers having very variable characteristics. Certain workers have obtained pulse measurements which indicate the permanent existence of a series of thin layers of nearly constant height and electron density, extending from 1 km. upwards to about 15 km. The existence of these layers has been deduced entirely from pulse measurements, but at these heights direct observation of the electrical state of the atmosphere can be made by means of balloons, and

observations do not reveal such ionised regions. There are also theoretical considerations which make it difficult to accept the existence of layers having the characteristics suggested, and the interpretation of these pulse results is, therefore, a matter of doubt.

The actual values of electron density in the various layers depends not only upon diurnal and seasonal changes, but upon the eleven-year solar cycle, which also affects the earth's magnetic field. During the years when sun-spots are most numerous, the electron density becomes greater. The last maximum year was 1937–1938, and we shall therefore reach a minimum condition in about 1943.

It has been noted already that the condition of the ionosphere has only been investigated over a few places on the earth. These are widely separated, however, and appear to show that the condition of the E and F_1 layers is much the same all over the earth (when sunlight conditions are the same), but there is more difference in the condition of the F_2 layer.

The results of continuous pulse-measurements made at Washington are summarised each month in the P.I.R.E. and a prediction made of the maximum usuable frequencies for the succeeding month.

We are now in a position to correlate the general facts of propagation theory for communication over great distances. We have shown that the three main layers of the ionosphere, namely E, F_1, and F_2, vary very considerably with time and season, and their condition will determine which has the maximum control on any frequency used for long distance communication. Before showing how the many variable factors which control communication may be co-ordinated into workable charts, or characteristics, to enable us to predict the correct wavelength to use for any given distance at any given time, we will briefly summarise the general features of long distance communication on the different wavelengths.

The gamut of the present useful wireless spectrum may be considered to extend from 30,000 metres down to a fraction of a metre, and as has been mentioned in Chapter I this spectrum can arbitrarily be divided thus :—

(1) The long and medium wave band = 30,000–400 metres.
(2) The critical wave band = 400–150 „
(3) The short wave band = 150–10 „
(4) The ultra short wave band = below 10 metres.

The Long Wave Band. Since both the earth and the *E* layer behave as good conductors to long and medium waves, their propagation is substantially represented as a "spherical transmission line" type. The wavelengths being commensurate with the height of the lower layer, the state of vertical electric polarisation with which long waves are usually emitted

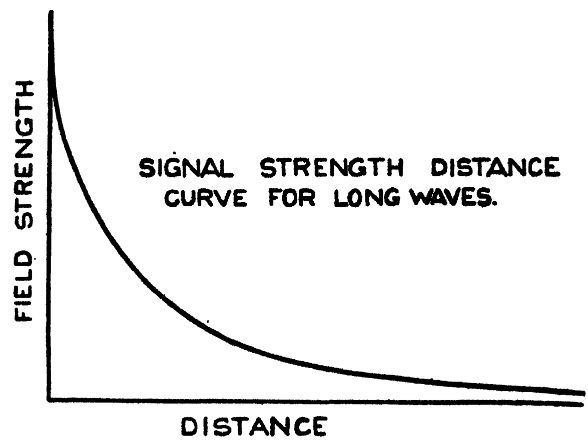


Figure 51.

persists (nearly) throughout their journey. Because the lower layer responds quickly to the sun's action, the day and night effects on long waves follow closely the actual coming of day or night.

Attenuation is the only factor which comes into the consideration of long wave propagation, and this will be proportional to $f^{\frac{1}{2}}$, hence the possible distance of communication with a given power is proportional to the square root of the wavelength used. With any given wavelength the signal strength distance curve is of exponential form as shown in Fig. 51, and the curve connecting wavelength and communication distance for a given radiated power will be as shown in Fig. 52. From the second curve it can be seen that only

waves above some 15,000 metres are suitable for communication over the greatest distances on earth, namely, half the circumference.

The Critical Wave Band. The critical wave band includes waves having frequencies near the resonance frequency discussed on page 86. As can be imagined the transition from long wave propagation phenomena to short waves is not sudden, but goes through a phase whose condition shows certain characteristics of both types. Waves coming in this class are not confined therefore to frequencies approximating to the collision frequency, but include waves lying in the region between 150 and 400 metres where reliable long distance communication is impossible.

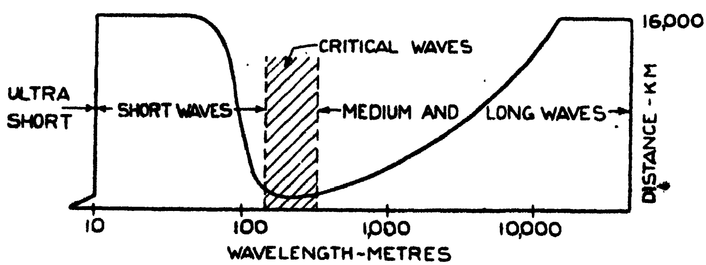


FIGURE 52.

During the daytime communication is confined to an area covered by the surface ray, which has a short range, and waves in this band are therefore suitable for short-range services, such as broadcasting, where the intention is to serve a limited area. During the dark hours, however, there is considerable reflection from the upper layer and hence signals may be transmitted over a considerably increased area. This longer night range is not necessarily useful because at points where the surface and reflected waves are of about the same amplitude considerable fading and distortion results, due to interference between them.

The Short Wave Band. Since short wavelengths are small compared with the height of the ionosphere, it is permissible to use the idea of rays as we do when discussing many problems in optics. It is necessary to remember, however, that the rays which we draw in explanatory diagrams have

no separate existence, and are merely representative of an indefinite number of other possible paths. Thus in Fig. 53 each line should be considered as representing a sheaf or pencil of contiguous rays assumed to spread as the ray is propagated outwards, and if the medium through which the ray pencil passes is homogeneous, the "illumination" of any section of the path is uniform. Short waves are able to penetrate the E layer because there is time for several oscillations of an electron at these frequencies before a collision with a gas molecule takes place, but this is the portion of the path in which the greatest attenuation occurs, because there are more air molecules per c.c. present at the lower altitude and hence more collisions. The rays then travel on to become bent at the F layers, where they also suffer some attenuation. If the frequency is below the critical frequency of the layer at the time and at the oblique angle at which the ray strikes the layer, the ray returns to earth, again passing through the first layer, but if the bending is not sufficiently great the ray escapes through the layer and is finally lost to the earth. The plane of polarisation with which the ray is emitted may not persist and in fact rarely does. Short waves generally will penetrate into the F_1 and F_2 layers, and since the electron densities of these layers do not have any simple relationship with the sun's altitude, variations in short

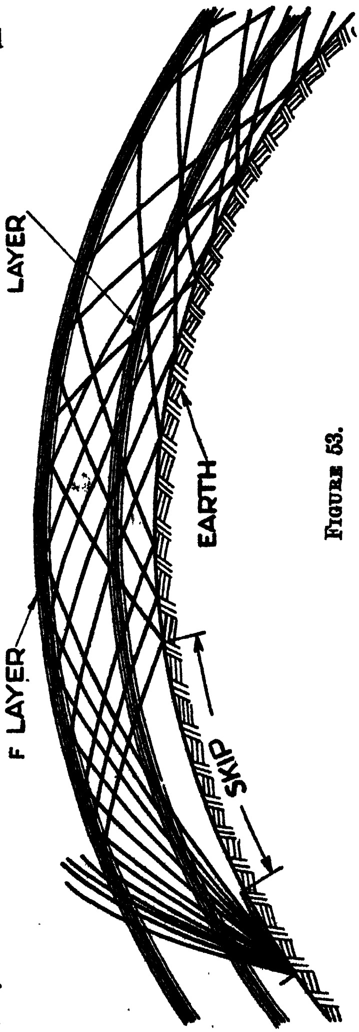


FIGURE 53.

wave conditions, whilst largely dependent on daylight and darkness, do not follow immediately and the connection is therefore less obvious.

The Skip Distance. Energy radiated horizontally from a transmitting aerial near the earth's surface is quickly absorbed due to the large earth losses, as has already been explained on page 67, and hence only short distance communication can be carried out by this horizontal radiation, which is usually known as the surface ray. Energy radiated at high angles may not be bent sufficiently at the layers to return to earth at all and therefore escapes. Shallow angle

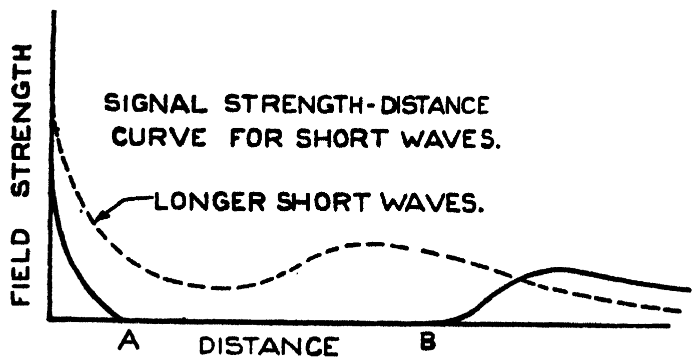


FIGURE 54.

radiation at an angle just great enough to escape absorption by the earth will, as shown in Fig. 53, enter the lower layer, suffer attenuation there, be bent at the upper layer and return to earth. A consideration of the figure indicates that between the distance at which the surface ray becomes negligible, and the distance at which the first ray returns to earth from the layer, there may be a zone which is not covered by any rays. This is called the skip zone and the distance across it the "skip distance." Signals are not necessarily absent in this area, as it will usually be "illuminated" by scattered radiation, to be defined later. Thus the signal strength-distance curve for short wave communication is of the form shown in Fig. 54, where *A* marks the limit of the surface ray, *B* the distance where the first down-coming ray is received and *AB* the skip distance, although it is more usual to consider

the distance from transmitter to B as the skip, because the range of the surface ray is always small. It should be mentioned that unusually short skip distances are occasionally recorded, due to an abnormal condition of E layer causing reflections from it.

In the case of the longer short waves (that is in the region of 100 metres) there may not be any definite skip distance but only a pronounced dip in the signal strength-distance

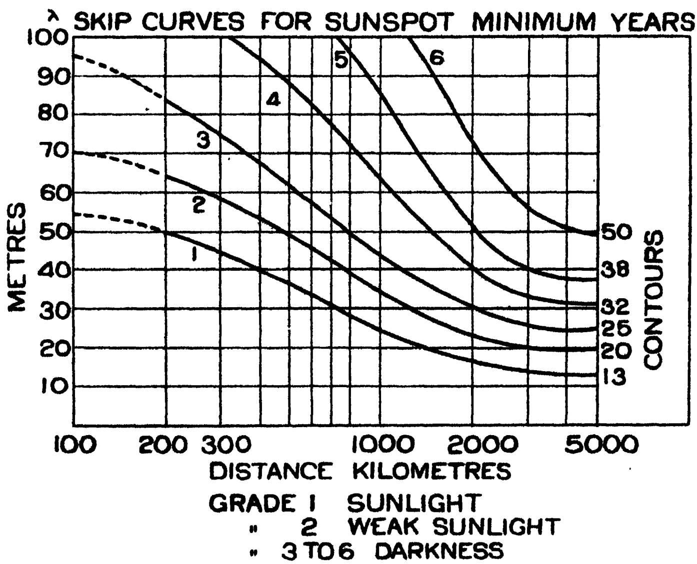


FIGURE 55.

curve, as shown in Fig. 54. With wavelengths in the neighbourhood of 12 metres, however, the skip distance may reach several thousand kms. when most of the path is in daylight, whilst if most of the path is in darkness no ray would return to the earth's surface at all. Thus the skip area will vary considerably in extent depending upon wavelength, season, and time of day, and Fig. 55 shows the skip distance for different wavelengths during daylight, twilight, and darkness conditions. The various grades, 1 to 6, are referred to later on in the chapter. These curves are really concerned with bending, and show the minimum distance at which the bending of a ray is sufficient to return it to earth, and it follows that for a

given distance and wavelength the darker the grade the less the bending.

It is to be noted that during the sunspot maximum years the skip curves are lowered by some 25% so that the limiting wavelength for sunlight conditions is reduced to some 9 metres.

Scattering. Within the *E* layer there are local centres or "clouds" of high electron density. Waves incident upon these may be scattered in all directions and hence may provide a very uncertain signal at a receiver. Scattering is more in evidence in the skip area because the only energy received may be due to this cause, but it is nearly always the means by which some of the energy reaches a receiver.

If the transmission is non-directional then scattering will take place from many points situated all round the transmitter. It follows that a direction-finder placed where the energy is mainly received by scattering will not indicate a bearing, since energy is reaching it from many directions.

It has been shown by T. L. Eckersley that in some cases where directional transmission is taking place, a direction finder placed within the skip zone will indicate the bearing of the station as being more or less where the "beam" enters the layer and not the true position of the station, thus showing a marked scattering of rays from this point.

Exactly how the earth is "illuminated" beyond the skip distance by rays bent back from the upper ionised layer has been a matter for a number of theories from time to time, but the generally accepted one is that the rays continue in a series of ricochets, that is by successive reflections from the earth and refractions from the upper layer, as shown pictorially in Fig. 53. If this is so it will be seen that the earth beyond the skip zone will be illuminated by a whole series of rays, from very shallow-angle rays nearly tangential to the earth's surface which have proceeded by a few ricochets, to rays near the critical angle which have "hopped" many times. The high-angle rays which make many ricochets will, in general, be more attenuated in traversing a given distance over the earth's surface than the low angle rays, because each reflection at the earth involves loss and each passage through the layer also involves attenuation. Thus it will be seen that the most effective rays for long distance transmission are those that leave the trans-

mitting aerial at a shallow angle, and hence an efficient aerial system should concentrate the radiation between about 3° and 15°.

During certain conditions of the ionosphere it appears that a type of transmission is possible which omits intervening ricochets, so that if the initial and final conditions of layer are correct for bending, the intervening condition need only confine the ray to the ionosphere and not necessarily return it to earth.

Reflection from the earth's surface has been discussed on page 60, and it will be seen that at a certain critical angle there may be practically no reflected, vertically-polarised ray.

In order to communicate successfully over a long distance a consideration of Fig. 54 shows that in addition to the attenuation not being too great to permit of the reception of sufficient strength, it is necessary also that the receiving station should not be within the skip zone. This means that the bending in the upper layer must be sufficient to bring down rays within the required distance. The attenuation and the bending of a ray of a given frequency are dependent upon the state of ionisation of the layers, and this in turn depends on the day-night conditions. We shall expect therefore that to maintain a service it will, in general, be necessary to alter the wavelength used to suit the time of day and season, remembering that the wavelength chosen will generally be that which gives least attenuation with sufficient bending.

Attenuation. This is now, not a function of the losses in an equivalent transmission line as it is for long waves, but it is a question of the absorption of a wave as it passes through the ionised layer, and it is found that attenuation is proportional (for daylight conditions anyway) to the square of the wavelength ; in fact exactly the reverse of long-wave attenuation. If the distance of communication depended only on attenuation, the curve for communication distance against wavelength would rise from the critical wavelength as the wavelength decreased. But as the bending effect, which will be considered next, is equally important, we reach a limiting value of shortness of wave for communication on earth, below which the range suddenly falls off to the surface-ray range. This limit is shown on the left-hand side of Fig. 52, where an abrupt drop

in the curve is shown at the wavelength of about 10 metres.

This does not mean necessarily that 10 metres will be chosen for communication to the antipodes; for one thing it is too near the critical value, and secondly, half the earth is never all in intense daylight (when the bending is greatest) and thus one rarely finds in practice any long distance communication carried on at wavelengths below 14 metres.

The amount of attenuation is dependent upon the layer conditions and in general will be proportional to the amount of ionisation. Thus during the day, because ionisation is greatest, attenuation will be greatest, whilst during the night attenuation on all waves is much reduced. For the same reason the attenuation in summer on any given wave will be greater than in winter. Where the route is mainly in daylight, the greatest attenuation will be observed when the ionisation is greatest at a point half way between transmitter and receiver, this occurring about one hour after midday (at that place).

Bending. The bending experienced by a ray of given wavelength depends upon the total change of electron density through which it has passed. The amount of bending with the same change of density varies as the wavelength so that waves below 10 metres are rarely sufficiently bent to return to earth even when the layer is in its most highly ionised condition.

The ionisation increases quickly after dawn and the amount of bending therefore increases considerably, and on any given wavelength, is greatest during the day, gets slowly less during the evening and early night, and is least during the late night period preceding the dawn. Hence the skip zone will be greatest for any given wavelength during the late night and least during the middle of the day. For the same reasons the skip will be greatest during the winter at any given place and be greatest in places of high latitude. Unlike attenuation, however, the condition of layer at receiver or transmitter is more important to bending than the condition between, because the bending into and out of the layer is chiefly concerned with the layer condition at the ends of the routes. Further, that end in the darker grade is really the controlling factor in this respect, for it is of no use a ray being

bent into the layer if at the other end there is not enough bending to return it, or vice versa.

The Zenithal Angle. The zenithal angle is of considerable importance, quite apart from the fact that high angle radiation is lost because it is not returned to earth. The reason for this is that a very narrow pencil of radiation is sufficient to illuminate the whole earth beyond the "skip" by successive "ricochets," and radiation at other angles only leads to the production of "echo" signals. This is shown in Fig. 13, where *H*, *B* and *C* are three rays leaving the transmitter at different angles such that successive reflections, one of ray *H*, two of ray *B*, and three of ray *C* bring the rays together at a receiver placed at *X*. The different lengths of path of these rays causes fading and blurring of the original signal.

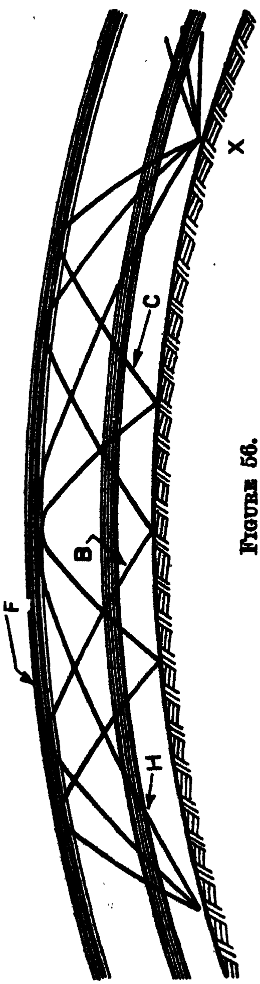


FIGURE 56.

The particular angle which gives the best results at great distances is a matter for consideration and will vary with different circuits, but it is certain that high-angle radiation is of no value and there is strong evidence that the most useful energy is that which leaves the aerial at a very shallow angle, just sufficient to avoid earth losses. Angles approximating to 8° are found best for long distance work, increasing to about 20° for nearer services. Hence the tendency in modern design of short wave aerials is to avoid those forms which give high angle-radiation, such as the harmonic aerial, and to provide systems which propagate energy at a shallow angle, such as the Franklin multiple aerial, which will be mentioned later.

The Ultra-Short Wave Band. We have already stated that ultra-short waves are only suitable for short-distance

communication and the division between short and ultra-short waves is usually made at 10 metres (30 Mc/s) because reliable long-distance communication is not possible below this wavelength. As has been explained, this is because there is insufficient bending even over the longest all-daylight path possible on the earth's surface. It is evident, however, that because the degree of ionisation varies so much from place to place and time to time that the minimum wavelength value which will be bent back to earth is not clearly defined. Thus although the London Television transmissions on 6·67 and 7·25 metres are only of value for direct-ray communication, occasional strong though usually distorted reception is obtained over very great distances, as, for example, South Africa and America.

An extended series of observations have been made at the Riverhead Station of the R.C.A. Communications Inc.[29] on European ultra-short wave television stations. The afternoon sound transmission from London was received on most days between September and March and the vision less frequently. The strength of the sound transmission sometimes exceeded 40 db. above 1 microvolt per metre. Reception is evidently dependent upon sufficient density in the F_2 layer and the reception conditions appear to follow fairly closely the critical-frequency for the F_2 layer, and when the critical-frequency dropped suddenly in January to low values, signals were not usually received.

Conklin[28] has collected data of amateur 56 Mc/s transmissions over distances of 400 to 1,200 miles, it having been found that communication over these distances is frequently possible. These successful receptions appear to depend upon the sudden increases in the electron density of the E layer mentioned previously.

Factors Influencing the Choice of Wavelength. It should now be clear that, broadly speaking, we shall choose for a given communication circuit the shortest wavelength which will ensure rays being sufficiently bent to return to earth within the required distance, since the shorter the wavelength the less the attenuation. As a general rule then, the shortest waves (in the neighbourhood of 15 metres) will be used to cover long distances which are mainly in daylight and longer waves

for darkness conditions. The problem is complicated by the fact that many long distance transmission paths are partly in darkness and partly in daylight, and hence the wavelength used must be a compromise. Further, since conditions are continually changing, more than one wave may be necessary to maintain any particular service throughout the 24 hours.

From what has been said, it is clear that the prediction of the performance of a short wave circuit, or the choice of a suitable wavelength for a given service, is a very difficult matter. A variety of charts have been produced from time to time, to enable the information to be obtained for average conditions, and these charts generally take one of three forms.

(1) The ionosphere conditions for a given number of seasonal conditions and for all latitudes may be charted, so that with the correct season-chart and a complementary map of the world, it is possible to observe the conditions over any route and at any time. The chart showing the ionosphere conditions will usually be made transparent, so that by sliding this over the map to the correct position for the time of day, the general ionosphere condition on the great circle route between the two places can be observed. Then by reference to "skip" curves and charts for field strengths for the different layer conditions, it is possible to estimate the probable field strength for a given power and distance. Such a system is, of course, universal in its application, but experience is needed for correctly interpreting the information so derived for those routes which pass through several grades of electrons density in the ionosphere.

(2) When one is concerned only with transmission from (or reception to) a given place, it is possible to produce a series of charts to show the best wavelength to use at any time and season, for communication with a place at any latitude and at any given distance. Such a set of curves is simple to use, but as they refer to only one place, they are not universal in application, and to make them so would necessitate the production of a very large number of sets for the different latitudes.

(3) It is possible to produce a comparatively simple series of curves for giving the field strength between stations of given latitudes and for a limited range of distances, not too great or small.

We propose to indicate the basis on which such sets of curves are produced, and in Appendix II will be found a series of curves suitable for general work.

(1) Eckersley and Tremellen Charts. Although recent charts are produced in a different form and will be described later, the charts as originally produced were in a pictorial form which lend themselves better to a general explanation of the principles on which charts of this type are based.

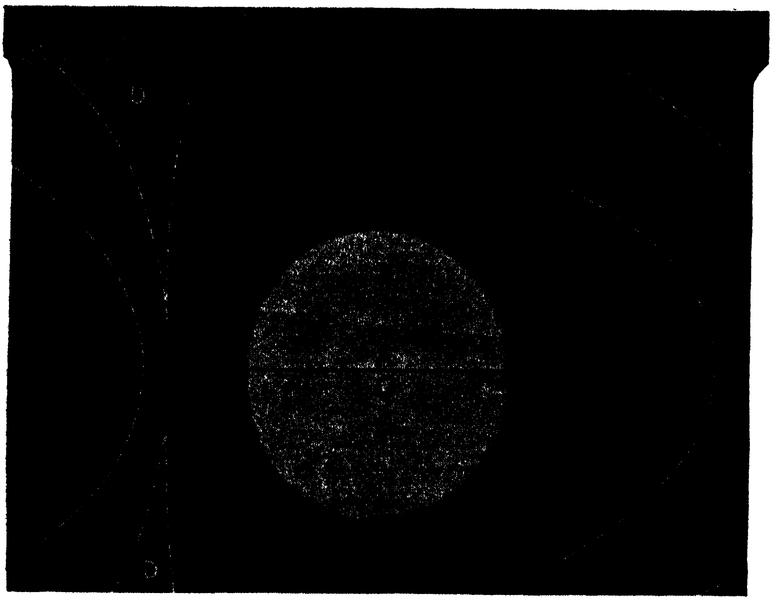


FIGURE 57.—EQUINOX.

Three charts, to cover equinox, summer and winter seasons, are shown in Figs. 57, 58 and 59. In these charts it will be observed that the infinite shades of ionosphere conditions have been limited to six grades, *A*, *B*, *C*, *C*–*C*¹, *C*¹ and D, these grades corresponding to: (*A*), conditions of intense sunlight; (*B*), weak sunlight; (*C*–*C*¹, *C*¹), twilight; and (*D*), darkness. These various grades are shown labelled in the Figs. 57, 58 and 59 and these grades *A*, *B*, *C*, *C*–*C*¹, *C*¹ and *D* correspond to the six grades 1 to 6 in the skip curves of Fig. 55. The sunset–sunrise line is also indicated by a thick black line,

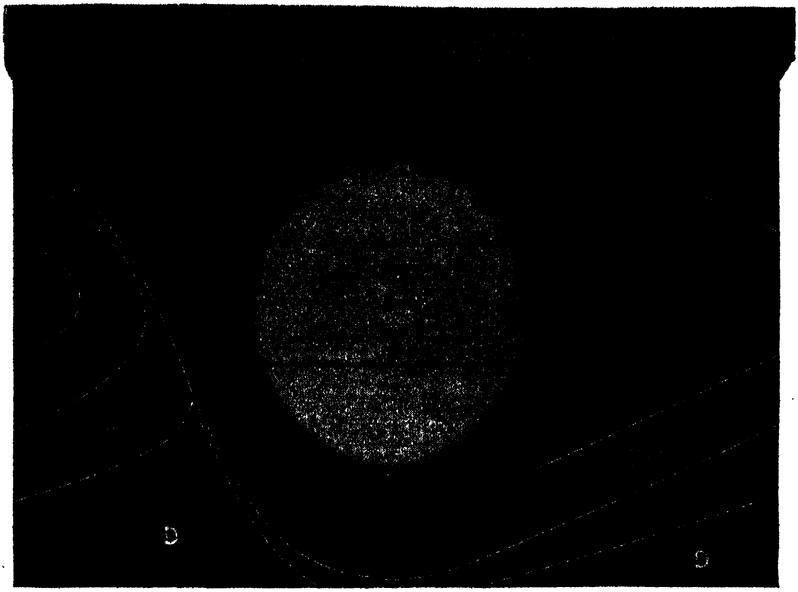


FIGURE 58.—SUMMER.

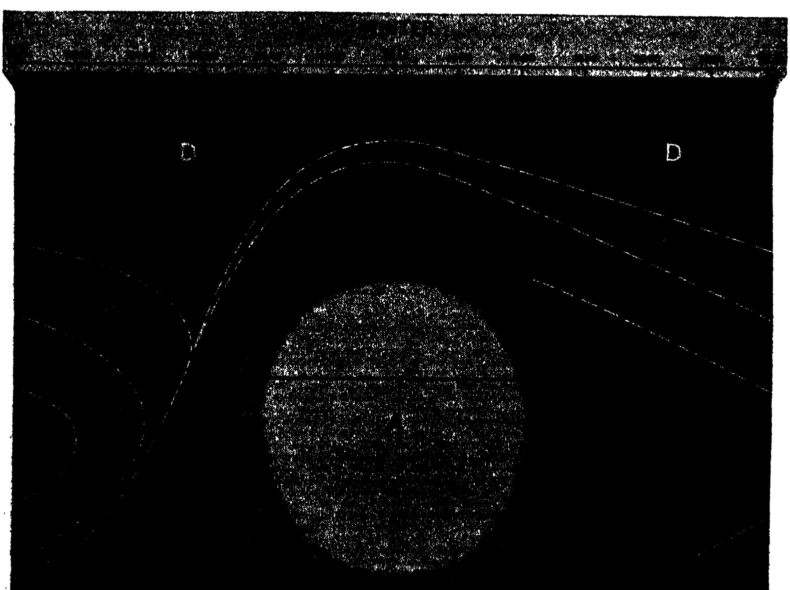


FIGURE 59.—WINTER.

which takes the form of two vertical lines at 6 a.m. and 6 p.m. at the equinox period and of "cosine-shape" curves during winter and summer.

The fact that—although the E and F_1 layers follow approximately the sun's altitude—the F_2 layer behaves in a contrary fashion, presents difficulties in the production of such charts, but the average behaviour suggests the necessity for the lagging effects of the darker grades in the charts as shown. It may help to explain the utility of such charts if the transmission characterictics of the various grades are enumerated.

Transmission Across the All Daylight Zone. The longest distance on the earth which can be said to be all in intense daylight at any one time is about 6,000 miles, and as the attenuation over such a daylight path will be high, the shortest possible wavelength will be used. This will be about 14 or 15 metres, as shorter waves than these would have a skip distance exceeding 6,000 miles. A shorter distance would necessitate an increase in wavelength with a corresponding increase in attenuation, and the great attenuation of this wave will prevent the communication extending much beyond this distance. The minimum signal will occur when it is a little past mid-day at a point halfway along the path.

Transmission Across the Twilight Zone. Across the twilight zone the attenuation on all wavelengths is much less than for a daylight zone and for wavelengths below 20 metres becomes especially small. When the seasonal condition and time of day is such that the great circle between the stations is near the dividing line between daylight and darkness (called the shadow line), the attenuation is so small that signals may make more than one journey round the earth, and produce what are termed "round the world echoes" at the receiver. For at this time the ionosphere condition along the great-circle path between the stations is the same around the whole earth, and hence it is easy to find a wavelength exactly suitable, whereas if the great circle line lies across different grades of layer condition, we cannot find a wave which will pass through these varying areas, for any wavelength which is long enough to pass through the darker grades will be rapidly attenuated in passing through the daylight zone. It might be thought that it is not possible for two stations to be in the all twilight

zone except for a brief period each day, but a reference to the charts will show that what is there termed "twilight" includes weak sunlight such as experienced in high latitudes in winter. For instance, during an English winter we are never in "daylight" within the meaning attached to it in the charts, and when working to other stations in the northern hemisphere the great circle line will be either in a twilight grade or a darker grade throughout the 24 hours.

We may say that waves suitable for routes in the twilight grade extend from 14 metres up to about 40 metres, the longer distance services utilising the shorter wavelengths.

Transmission Across the Darkness Path. Across this zone between the twilight and the late dark area the attenuation on wavelengths of the order of 20 to 60 metres is very slight, but on waves about 20 metres the signal strength would commence to fall off rapidly, not because of the attenuation, but because of insufficient ray bending.

Transmission Across the Late Darkness Zone. For the portion of the earth which may be considered as in late darkness, waves of greater than 50 metres are suitable, and 30 metres represents a critical wavelength below which bending will usually be insufficient. Waves less than 30 metres may pass successfully through a short length of late darkness path, if it is intermediate between the stations and not over one or the other. This to some extent can be understood, because, as previously mentioned, if the conditions at the commencement and the end of the ray path are sufficient to create a bending of the ray into and out of a path coincident with the layer, then the intermediate condition is concerned chiefly with attenuation and not bending.

The type of charts now being used are not divided into arbitary grades of darkness as just described, but in contour lines of critical wavelengths for oblique reflection. This is, of course, essentially the same, since (as we have already seen) there is a direct connection between electron density and critical-wavelength. The dividing of the ionosphere into such critical-wavelength contours, however, gives a direct indication of the minimum wavelength to which the layer is not transparent over any given great-circle path that is being considered.

(2) Field-Strength Contour Charts. An alternative method of producing charts has been suggested which aims to map contour-lines of field strength from any point of given latitude, for different wavelengths, seasons, and times of the day. Such a system would be an ideal one, but unfortunately a prohibitive number of charts is necessary, since, not only must a chart be provided for each hour of the day, and the various seasons, but also for every 10° of latitude, and, of course, for a large number of wavelengths. This would necessitate about 1,400 charts to cover the short wave-band.

Such a chart is shown in Fig. 60, and this is given as it is illustrative of the conditioning features of short wave transmission. The chart shows a series of contour lines of equal field-strengths from a 1 kW transmitter, situated in England (latitude 52°), the time being midday, in winter, the contours being denoted in db, plus, and minus from a zero constant of 1 μ V per metre. The small shaded area in the centre is the skip area, after which the ionosphere rays are returned to earth. From the skip area the rays follow a richochet course until they encounter the dark grade (shaded), which has not sufficient electron density to return them to earth again Thus from the skip area to this dark area we have contour lines of equal field strength, spreading outwards. Towards the south the rays pass through an all daylight grade where the attenuation is highest and is the conditioning feature. To the east and west the daylight is less intense and attenuation in consequence less, and in regions near the shadow-band (shown dotted) is very small indeed. This causes the contour lines to be drawn out towards the shadow line, so that high field strengths are obtained at very great distances from the transmitter.

To the N.E. and N.W. the direction is into the darkness area and the possible communication distance is thus shortened and, in fact, signals cease almost abruptly, owing to the lack of bending, because the region of critical density has been reached.

Charts for Field-Strength, Distance, for Different Latitudes. Yet another type of chart, which is useful for a limited range of distances only, has been developed, which gives the field-strength direct, for different seasons, and times of the day, and between stations of different latitudes. A series of such charts is given in Appendix II.

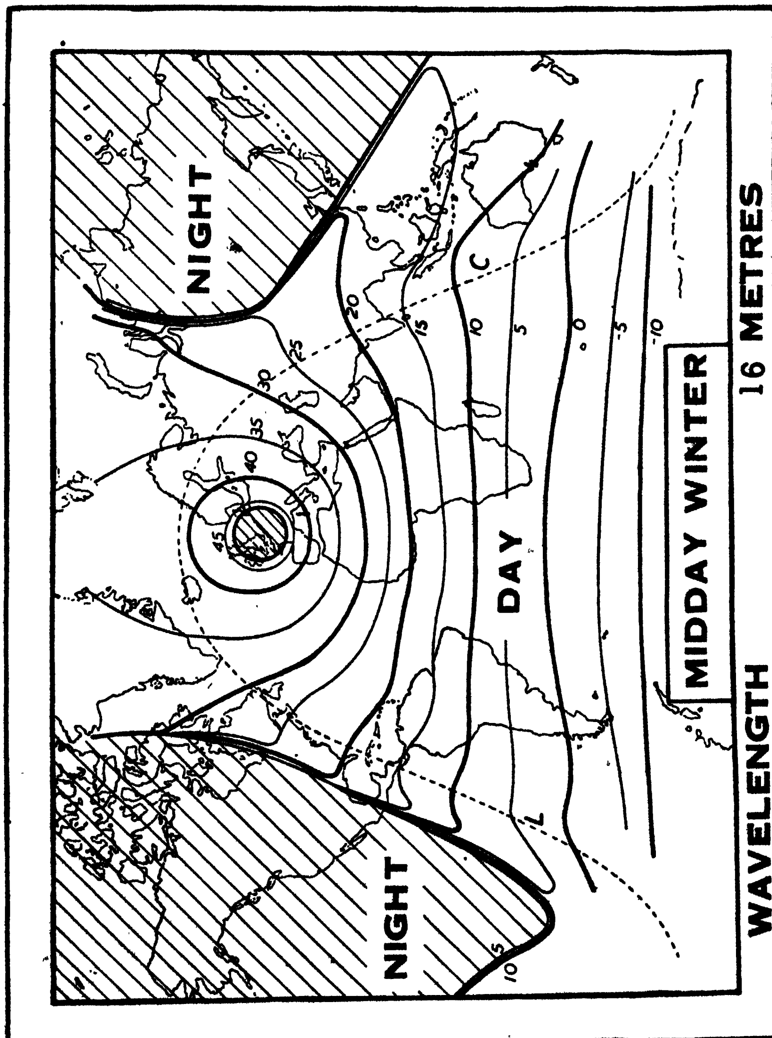


FIGURE 60.

It now remains to discuss some peculiarities of short wave propagation, the principal being the fading phenomena.

There are two distinct types of fading giving the following effects:

(1) Complete cessation of communication for many hours.
(2) Variations of signal strength, the change of which may be slow or very rapid in character.

Complete Fading. Complete fading is liable to occur if the transmission path lies in high latitudes, and particularly if it passes near the magnetic poles, when it may be so bad as to eliminate signals for several hours. In one or two cases signals have disappeared for as long as one or two days, but usually these periods will be of short duration, and there will often be only a general reduction of the signal level and not a complete cessation. These complete "fade-outs" occur during magnetic storm activity, and such are often associated with sun-spots, which have a cyclic variation of about 27 days, the time of the sun's rotation. Systematic observations of the correlation of fading with sun-spots and magnetic storms have been made by T. L. Eckersley over a number of years, the sun-spot activity being observed by the simple expedient of pointing a focussed telescope to the sun, holding a chart a few inches below the eyepiece at such a distance that the sun's image is two or three inches in diameter, and plotting on the chart the position of the sun-spots or sun-spot patches. These vary considerably in size and number, but all those of interest are quite large enough to be observed by this means. It is found that those spots which pass exactly through the centre-meridian of the sun may have the most effect, and this is to be expected if the spot is a hole in the sun's envelope. Secondly, the influence on wireless signals appears from one to three days after the sun-spot has passed the centre-meridian, so that whatever is causing the fading is some agent which travels at a much slower speed than light. A further point is that the size of the sun-spot is no criterion of the effect produced, and many periods of great sun-spot activity may pass without any marked effects on wireless circuits. Sun-spots and magnetic storms besides having a period of activity recurring each month have also a long

cyclic variation, the periods of peak activity recurring every 11 years, 1939 being the last. Again, stations in high latitudes are chiefly affected by magnetic storms and sun-spots, and therefore it is assumed the agent causing the disturbance is divided by the earth's magnetic field towards the magnetic poles.

The fading has been attributed to increased attenuation due to the greater ionisation. Extensive tests on ultra short waves were tried over a long distance circuit during these abnormal times on the assumption that if the ionisation is so much greater the bending might be sufficient to bring them down to the earth's surface within the required distance. Except for occasional loud signals, no useful results were obtained, and there is no method yet known of overcoming this type of fading.

It should be mentioned that in addition to causing fading, the magnetic storm activity affects generally the ionosphere, particularly at high latitudes, the level of ionisation increasing and decreasing with the magnetic storm activity. This means that the correct wavelength to choose for any channel in high latitudes will vary cyclicly with the magnetic storm era, and as an illustration we might mention that the Canadian circuit was forced to employ waves as long as 60 to 70 metres as a night wave during 1933, whereas it used 30 metres during the maximum magnetic-storm period of 1939.

Catastrophic Disturbances. In addition to complete fade-outs being caused by magnetic-storm activity, a rare disturbance has been observed which affects all daylight routes. This disturbance, which may last for periods of a few minutes up to half an hour, is not definitely associated with magnetic storms, but has somewhat similar effects (except that they are world wide). It has been observed that these disturbances are associated with hydrogen eruptions in the sun's chromosphere.

Rapid Fading. It is usually considered that the causes of this type of fading are ray interference and change of wave polarisation.

Consider ray interference. As was seen in a previous section, a long-distance, short wave signal is not a single ray, but may be made up of two or more rays arriving by different paths.

If we have only two rays, the phase of these will determine the resultant field at that point, which can vary from zero, if the rays are equal and completely out of phase, to their sum, if in phase. If more than two rays are involved the resultant depends upon their vector sum.

It is found that even if the receiver is in such a place that it cannot be energised by ricochet interference, fading can still occur. In this case the interference is between closely adjacent rays following approximately the same path. If the ionosphere were a stable medium merely graded vertically in a definite unchanging manner, then these contiguous rays would "illuminate" uniformly a small area of the earth's surface. As, however, the ionosphere is not homogeneous either horizontally or vertically, the different rays forming the ray-pencil are disarranged in their passage through it and now "illuminate" the area in a non-uniform fashion. As the ionosphere conditions vary from instant to instant, the distribution over the area considered is continually changing and hence fading of the signal results.

Actually the variation of signal strength is greater than a calculation giving the resultant of a number of vectors having random phase relationships would lead us to expect. This is to be accounted for by an additional variable factor, namely, changing polarisation of the waves. Thus if we use a vertical aerial for reception it means that only vertically-polarised waves, or those waves giving some vertical component, will be received. Hence a single wave, the polarisation of which changes with time, will produce a variable signal apart from any interference variation.

It is of interest to note that fading is different in time at points quite near together in space. Thus we can, to some extent, overcome fading in the following ways:

(*a*) By using both vertical and horizontal aerial systems.

(*b*) Summing the energy received on a number of aerials spaced apart a sufficient distance.

Very deep fading is frequently found when a station is near the edge of the skip distance. This is due to the fact that very small changes occurring in the ionosphere make considerable changes in the direction and intensity of the rays coming down

near the station, at one moment the rays falling between the transmitting and receiving stations, now passing over the latter.

At any one place the fading is different for different frequencies separated by only a few cycles. If fading is caused by interference between rays following different paths this will readily be understood, for the bending of a ray is a function of its wavelength and quite a small difference of wavelength (and therefore bending) will change the points at which a group of rays again meet. Thus a continuous-wave signal suffers more violent fading than a modulated wave because the latter involves the transmission of a wide spectrum (see Chapter III), and there is a better chance of collecting part of the signal in this case, because the individual frequencies traverse different paths. For this reason telegraph transmitters are frequently modulated, so that the energy is conveyed to the receiver by a band of frequencies instead of a continuous wave being used. Although a modulated carrier may suffer less fading, such a signal is often received distorted because the component waves are altered in relative value. This is sometimes referred to as selective fading. This does not matter for telegraphy, but may be serious for telephony or facsimile telegraphy. If the fading is due to a number of rays adding together at random phase, the distortion is found to be serious and when the distortion is small it is probably because the fading is chiefly due to changing polarisation. Yet another type is characterised by fading of the modulated component but no apparent fading of carrier wave. One of the authors has suggested* that this type of fading is due to a phase-shift of carrier relative to the side bands as the wave passes through the ionised layer, for it can be proved that** a phase shift of carrier of 90° relative to its correct phase can almost eliminate amplitude modulation although a frequency modulation is substituted. In short wave communication, fading is generally the most difficult feature which has to be overcome, as it is to be found at all times, on all wavelengths, and it is this feature which necessitates an average level of signal strength for commercial circuits much above the noise level of the circuit.

Modern technique is gradually overcoming the ordinary fading phenomena and an up-to-date receiver can deliver a

* Marconi Review, No. 23. ** See Chapter III.

constant level of signal at nearly all times, of a quality good enough for commercial telephony.

Echoes. It is found that when a single short wave signal is transmitted, more than one signal is sometimes picked up on a distant receiver. It has become customary to speak of all the received signals subsequent to the first as "echoes," though these signals are not produced in the same way as an ordinary echo, i.e. by a wave reflected from a large object and returning back more or less along the path of the incident wave.

Although echoes have actually been heard on the longer waves, it was the short waves which first called attention to their existence in wireless telegraphy, and in general they may be classed under three headings:

(1) Very long delay echoes.
(2) 1/7 second echoes.
(3) Quick echoes.

(1) **Very Long-Delay Echoes.** Long-delay echoes called the Stormer echoes, after Carl Stormer who first observed them, may appear as long after the signal as 10 seconds (30 seconds has been mentioned) very much distorted. In fact it is because the echo is so unlike the original signal and so long after it that its existence was unobserved for a very long time after short waves had been a common means of communication. Naturally an echo with such a delay is very difficult to account for, as at the velocity of light the echo signal would have to travel hundreds of times round the earth, and theories put forward are various in number. There is no scope within the present book to set out these theories, particularly as the long-delay echo is not of great interest to the wireless engineer, and none of the theories put forward so far have been established. It should be pointed out that the Stormer echo is such an extremely rare phenomenon that it has only been reported by a few observers in Europe and Asia.

(2) **1/7 Second Echoes.** These echoes, the "all-round-the-world" echoes, are caused by a signal making a complete circuit of the earth. Thus a station receiving from the west, say, may get an echo signal from the same direction approximately 1/7 second later than the original signal, should the condition of the layer be favourable. Although weaker than

the original signal, it may be at times troublesome, as a directional system cannot eliminate it. The times at which these echoes are heard usually coincide with the period of season and day when the great circle line between the stations is nearly coincident with the shadow band, and wavelengths between 15 and 18 metres show this type of echo in a very pronounced manner.

"All-round-the-world" echo is a common phenomenon on a local station at the right season and time, but a more rare phenomenon on long-distance working. This will be understood by studying the shadow charts; for a local transmitting station is in the same grade as the receiving station and at periods when the stations are near the shadow band the signals will follow the shadow band great circle path. On the other hand if a long-distance station is considered, there is only a short interval of time at one particular season which can possibly create echo; further, the attenuation of the signal along the extra distance to be covered between transmitter and receiver makes the chances of echo less.

In addition to the "all-round-the-world" echo, a shorter-interval echo will often be heard at the same time, which is caused by a portion of the transmitted energy taking an opposite path round the earth and so arriving at the receiving station from a direction approximately opposite to that of the main signal and at a time interval dependent upon the positions of transmitter and receiver. This echo is not troublesome if directional systems are used as they will usually eliminate it or render it too weak to have any serious effect. These two echoes are always clearly defined and there is no appreciable distortion

(3) **Quick Echoes. (Multiple.)** Quick echo is made evident in different ways; thus when a tone or speech is being received, a blurring or distortion of the signal may result or it may have a "hollow" sound giving the ordinary "empty room" effect. Signal blurring and distortion are due to multiple ray reception and as the distance increases their effect becomes less marked. This is explained by the fact that the rays making the longest "ricochets" are less attenuated, for at each ricochet from earth and at each passage through the E layer a little energy will be lost and thus at very great

distances there may be only one really strong ray left, in consequence of which the echo signal will be negligible. This quick or multiple echo is so close behind the main signal that when recording telegraph signals it shows up merely as a lengthening of the " marking " periods as if the relays had been given a " marking bias." For this reason the effect was known as " ether bias " in the early days of short wave working, when

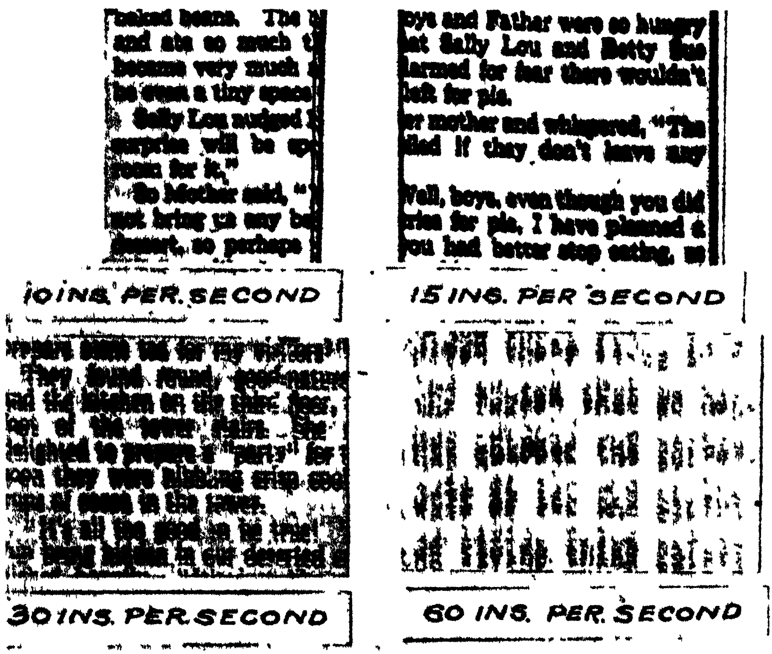


FIGURE 61.

the reason for it was not clearly understood. Except that it may cause fading, this multiple echo does not interfere with telegraph reception, as it may be corrected by the use of suitable " shaping circuits." The introduction of facsimile showed the effect up in its true light, namely as a series of separate signals, an example of multiple in facsimile being shown in Fig. 61, and measurements since made show that the blurred signal is in reality a number of signals arriving by paths of different lengths, one milli-second being the greatest time interval yet observed between individual rays. By careful check of these

multiple echo times it is but a geometrical matter to arrive at the number and length of each ray path, from which an estimate can be made of the height of the ionosphere.

In addition to multiple rays giving the quick echo, scattering has also some part in the various multiple effects observed, for if the scattered radiation is present at a strength comparable to the main signal this slightly modifies the sound heard and the blurred signal is most probably due to both multiple and scattered signals combined. The "empty-room" effect which is unmistakable when heard, is generally considered to be due to scattering, the scattered signal providing a background to multiple signals which gives an effect very closely analogous to the sound background in an undraped empty room.

Atmospherics. It is customary to imagine that atmospherics are negligible on short waves. Actually they are quite strong on the longer short waves during the summer months, and it is only on waves below 15 metres that they can be said to be quite negligible. But on waves of this order we have considerable increase of general noise level due to machines, magnetos and loose mechanical contacts.

Atmospherics are, of course, electric disturbances in the lower air layers caused by lightning discharges. The effect of atmospherics on communication is not so much dependent upon the intensity of any single atmospheric, as on the frequency with which they occur. Because the power involved is so enormous, an atmospheric centre thousands of miles away can create serious interference. Thus wireless services conducted by stations in temperate climates can be more upset by atmospherics emanating from zones near the tropics than from local storms, because the latter are infrequent whereas the former are almost continuous.

Most of the atmospheric producing centres are large land areas near the tropics such as Africa, Northern Australia, Northern South America and India. These atmospheric centres do not remain stationary but vary periodically with the sun, moving some 10° north during our summer and south some 10° during our winter. In addition to the seasonal change of position of the atmospheric centres, the actual amplitude and frequency, particularly frequency, vary with

the sun's altitude diurnally, their frequency reaching a maximum at 3 p.m. local time.

The atmospheric pulses produced are very varied in character, but since their wavefronts are usually steep they are capable of giving interference on an infinite spectrum of waves (since a pulse is an infinite series of periodic frequencies), and thus the atmospheric is to be treated as an interfering transmitter radiating power at all frequencies.

At 100 metres, atmospheric interference is extremely bad in tropical zones, and down to 30 metres the interference can be serious at times, at all places. It is to be observed that stations in a temperate zone may suffer more from atmospherics on the very short waves than stations in the tropics very close to the atmospheric centre, for the tropical station will be inside the "skip" area and thereby unaffected.

Measurement of Received Signal Strength. In the foregoing sections an outline has been presented, making use of various theories and assumptions regarding the ionosphere, etc., and certain methods of investigating ionosphere properties have been given. We will now discuss one or two other measurements associated with the subject.

One of the most important and practically useful measurements in connection with any investigation into the propagation of wireless waves must be that of received signal strength. Even on long waves this presents considerable instrumental difficulties if weak signals are to be measured, because the incoming power is insufficient to measure by direct means.

When measuring short wave signals grave difficulties are introduced by the nature of the signal, since this is rapidly varying in strength and is usually polarised in a complex fashion. In addition, the high frequencies involved make the design of accurate apparatus very difficult, as extremely small accidental capacity couplings may so greatly modify results.

Practically all signal strength measuring apparatus depends upon matching the signal against a known output from a local source, using the same receiver for both signal and local source, one of the most important points in design being the careful screening of the latter. It is usual to employ, for reception, a frame aerial, as by this means the magnetic field in all directions can be measured and also the effective

area (which must be known in order to express results in microvolts per metre) can be determined better than the effective height of an open aerial.

Measurement of Received Wave Polarisation. Appleton and Ratcliffe examined the downcoming wave of a south to north transmission at broadcast frequencies and found it to be circularly polarised with a counter-clockwise sense of rotation, whilst a corresponding experiment conducted in Australia showed a clockwise rotation, thus demonstrating that the circular polarisation is produced by the earth's magnetic field. Eckersley has obtained simultaneously oscillograms of the same signal received on a horizontal and a vertical aerial, and found that in some cases the fading on the two is opposite in phase, thus indicating an elliptically polarised wave.

The cathode-ray direction finder of the Radio Research Board has been employed to examine the polarisation of short wave reception. The signals from two loops at right angles to each other are put through identical receivers (adjusted to give the same performance) and applied to the two sets of deflecting plates in a cathode-ray oscillograph. The arrangement is therefore similar in theory to a Bellini-Tosi direction-finder with the cathode-ray oscillograph as the goniometer, and a vertically polarised surface wave such as is produced in long wave transmission would give a straight line at an angle which would indicate the bearing of the transmitter.

When the arrangement is applied to short wave signals, however, an ellipse is usually produced, due to the presence of horizontally polarised components, though a rotating straight line is sometimes observed. The ellipse is continually varying both in direction of axes, in shape and in size.

It is evident that further extended study of the nature of the received signal at long distances is of considerable practical importance owing to its probable influence on the design of receiving aerial systems.

Direction-Finding. Although it might appear out of place to include mention of direction-finding in the present Chapter, the fact that one method used is an extension of the pulse transmissions for ionosphere measurements must be our justification.

A suitably designed direction-finder can obtain a bearing of an ordinary short wave transmitter only if the direct ray is large compared with the scattered radiation from the ionosphere, hence the satisfactory working distance is very small, signals from moderate distances, so necessary for direction-finding work, giving no bearing. If, however, the transmitter is caused to produce very short pulses and the receiver is equipped with a cathode ray oscillograph, then the direct ray stands out separately on the screen and may be used to take a bearing which is unaffected by the scattered radiation. All that is necessary is that the direct ray shall be in itself of sufficient strength to be utilised and not that it shall be much stronger than the scattered radiation and, as a matter of fact, good bearings can be obtained when it is much weaker.

It is proposed to use this method of direction-fin[illegible] connection with aircraft routes, particularly in tropic[illegible] tries where atmospherics make the usual medium wave apparatus very difficult to work. In the case of transmission from an airplane the range of the direct ray is considerable.

SELECTED REFERENCES

(1) *Propagation of Radio Waves Along Surface of Earth and in Atmosphere.* P. O. Pedersen. Published in English by G. E. C. Gad, Copenhagen, 1928.

(2) *High-Frequency Alternating Currents.* McIlwain and Brainerd. Chapman and Hall.

(3) *Spreading of Electro-Magnetic Waves from a Hertzian Dipole.* Ratcliffe, Vedy and Wilkins. *J.I.E.E.*, Vol. 70; May, 1932.

(4) *Some Measurements of the Electrical Constants of the Ground.* Barfield. *J.I.E.E.*, Vol. 75; August, 1934.

(5) *Electrical Measurements on Soil with Alternating Currents.* Smith-Rose. *J.I.E.E.*, Vol. 75; August, 1934.

(6) *Propagation of Radiation Having Wavelengths of 2 and 3 Metres.* McPetrie and Saxton. *J.I.E.E.*, Vol. 87; August, 1940.

(7) *Propagation of Radio Waves over Surface of Earth and in Upper Atmosphere.* Norton. *P.I.R.E.*, Vol. 24; October, 1936; Vol. 25; September, 1937.

(8) *Reflection Curves and Propagation Characteristics of Radio Waves Along the Earth's Surface.* McPetrie and Strickland. *J.I.E.E.*, Vol. 87 ; August, 1940.

(9) *Ultra-Short Wave Refraction and Diffraction.* T. L. Eckersley. *J.I.E.E.*, Vol. 80, March, 1937.

(10) *Propagation of Wavelengths Between Three and Eight Metres.* Jones. *P.I.R.E.*, Vol. 21 ; March, 1933.

(11) *Random Fading of 50 Mc/s Signals.* Maclean and Wickizer. *P.I.R.E.*, Vol. 27 ; August, 1939.

(12) *Stability of Two-Metre Waves.* Burrows, Decino and Hunt. *P.I.R.E.*, Vol. 26 ; May, 1938.

(13) *U.H.F. Wide-Band Propagation.* George. *P.I.R.E.*, Vol. 27 ; January, 1939.

[illegible]tra-Short Wave Refraction. Smith-Rose and McPetrie. [illegible]*E.*, Vol. 11 ; January, 1934.

[illegible]port of C.C.I.R. Committee on Radio Wave Propagation. *P.I.R.E.*, Vol. 26 ; October, 1938.

(16) *London Television Service.* MacNamara and Birkinshaw. *J.I.E.E.*, Vol. 83 ; December, 1938.

(17) *Wireless Section, Chairman's Address.* Gill. *J.I.E.E.*, Vol. 84 ; February, 1939.

(18) *Application of Frequencies above 30,000 kc/s to Communication Problems.* Beverage, Peterson and Hansell. *P.I.E.E.*, Vol. 19 ; August, 1931.

(19) *Ultra-Short Wave Propagation.* Schelling, Burrows and Fevrell. *P.I.R.E.*, Vol. 21 ; March, 1933.

(20) *Ultra-Short Wave Phenomena.* Englund, Crawford and Mumford. *P.I.R.E.*, Vol. 21 ; March, 1933.

(21) *Propagation of Waves Below 10 Metres.* Trevor and Carter. *P.I.R.E.*, Vol. 21 ; March, 1933.

(22) *Radiotelegraphy and Telephony on Half-Metre Waves.* Uda. *P.I.R.E.*, Vol. 18 ; June, 1930.

(23) *Micro-Ray Communication.* Macpherson and Ullrich. *J.I.E.E.*, Vol. 78 ; June, 1936.

(24) 75 *Cm. Radio-Communication Tests.* Hershberger. *P.I.R.E.*, Vol. 22 ; July, 1934.

(25) *Micro-Radio Waves.* (Summary of Lecture.) Marconi. *Electrician*, Vol. 110 ; January, 1933.

(26) *The Magnetic-Ionic Theory.* Ratcliffe. *W.E.*, Vol. 10 ; July, 1933.

(27) *Wireless Studies of the Ionosphere.* Appleton. *J.I.E.E.*, Vol. 71; October, 1932.

(28) 56-*Megacycle Reflections by Sporadic E-Layer Refractions.* Conklin. *P.I.R.E.*, Vol. 27; January, 1939.

(29) *Observations on Sky-Wave Transmissions above* 40 *Mc/s.* Goddard. *P.I.R.E.*, Vol. 27; January, 1939.

(30) *The Sun and the Ionosphere.* Chapman. *J.I.E.E.*, Vol. 88; November, 1941.

CHAPTER VI

HIGH FREQUENCY FEEDERS

IT is the usual practice nowadays to install several transmitters (or receivers) in one building and erect the aerials or aerial arrays for them some distance away. Hence a special form of non-radiating link, usually termed a feeder, becomes necessary between the aerial and transmitter (or receiver) the function of which is to convey high frequency power with the minimum of attenuation. In the case of an aerial array, the feeder system will also have to be arranged to supply currents having the correct phase relationships to the individual aerials. Feeders take the form either of parallel wires or concentric tubular conductors.

The telephone engineer will at once recognise the feeder as a transmission line which is in some ways simpler than those with which he usually has to deal, but the coming of the feeder for high-frequency power transference introduced the wireless engineer to somewhat unfamiliar notions, and hence we shall endeavour to deal rather fully with its operation.

Adjustment of Load Impedance for Maximum Power. It may be useful here to deal with an important general condition to be realised in all telecommunication circuits if we are to get the greatest possible output from them, namely, the principal of matched impedance. Suppose that we have a resistive load R_o, supplied from an alternator, having an internal resistance R_g and internal reactance X_g. (Fig. 62.)

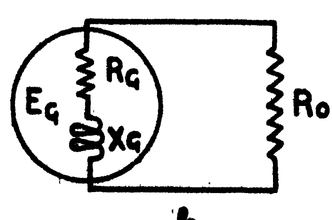


FIGURE 62.

The total impedance is

$$\sqrt{(R_g+R_o)^2+X_g^2}$$

and hence current I is equal to $\dfrac{E_g}{\sqrt{(R_g+R_o)^2+X_g^2}}$

and output W is $\dfrac{E_g^2R_o}{(R_g+R_o)^2+X_g^2}$ . . . . . (1)

The value of R_o which makes W a maximum will be that which makes $\dfrac{dW}{dR_o}$ equal zero.

Differentiating W by the quotient rule we have

$$\frac{dW}{dR_o}=\frac{(R_g^2+2R_gR_o+R_o^2+X_g^2)\,E_g^2-E_g^2R_o\,(2R_g+2R_o)}{(R_g^2+2R_gR_o+R_o^2+X_g^2)^2} \qquad (2)$$

Hence W will be a maximum for that value of R_o which makes

$$R_g^2+2R_gR_o+R_o^2+X_g^2-2R_oR_g-2R_o^2=0$$

from which $R_o=\sqrt{R_g^2+X_g^2}$ . . . . . . (3)

It will be seen that the maximum power is transferred to the load, at moderate efficiency, when its resistance is equal to the internal impedance of the alternator or other source and the load and source are then said to be matched.

This condition of matched impedance is continually being met with throughout telecommunications, whether generator or load are in close proximity, or whether they are separated by a feeder line. The only difference is that in the former case matching of load to generator automatically ensures maximum output, whereas in the latter case we have to match load to line to minimise line losses, and load and line to generator to obtain maximum output from the latter.

In many cases the source and load impedances will be fixed from other considerations, and it will then be necessary to insert some transforming device. In the case of voice-frequency work an ordinary transformer would be used with a suitable ratio of primary to secondary turns, but a transformer of conventional type would be very difficult to design for high frequencies, due to the magnitude of stray magnetic leakage and capacity effects, and, moreover, as will be seen later simpler solutions are available.

Electro-Magnetic Waves along Feeders. Fig. 63 gives a picture of what we may suppose is occurring when a feeder is connected to a generator. The dielectric between the two conductors will be traversed by electric strain lines stretching between the two conductors. Since these lines are in motion, magnetic lines will be produced, and these will be concentric with the conductors. In the "skin" of the conductors there will be a drift of electrons, or, in conventional terms, a current. No energy is required to maintain an electro-magnetic wave in a pure dielectric, but the current in the guiding conductors results in some energy being converted into heat (as in all electric conduction) and, in addition, there must be some loss in the solid insulation necessary to support the conductors. This loss results in a reduction of the velocity with which the wave travels down the

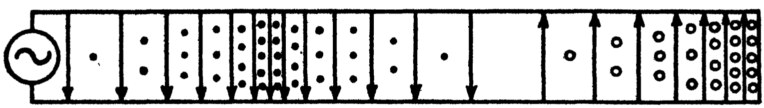


FIGURE 63.

feeder, whereas if there was no loss the "guided" wave would travel at the same speed as a "free" wave, i.e. at very nearly 3×10^{10} cms. per sec. In the case of feeders carrying very high frequency currents, the energy loss is fairly small compared with other effects and the reduction of velocity is small. Hence as a fair approximation a working theory can be developed neglecting loss.

Correct Termination to Prevent Reflection. Assume now that the wave reaches the end of a feeder which is open-circuited (or short-circuited), then since there is no circuit to accept energy there will evidently be reflection, and a wave will travel back along the feeder towards the generator. Here there may be a further reflection, so that the actual distribution of electric and magnetic lines is the resultant of a number of waves travelling in both directions along the feeder.

Thus if F_1, F_2, F_3, etc., Fig. 64, represent successive forward waves, these will form a resultant forward wave, which is their vector sum. Similarly, the corresponding reflected

waves R_1, R_2, R_3, etc., form the resultant reflected wave R_s. Since each journey up and down the line entails small attenuation due to line and leakage losses, each succeeding vector is of smaller amplitude than the previous one, and hence the general form the vectors take is a spiral.

Generally speaking, the resultant waves formed will be of small amplitude, but should the length of line be such that

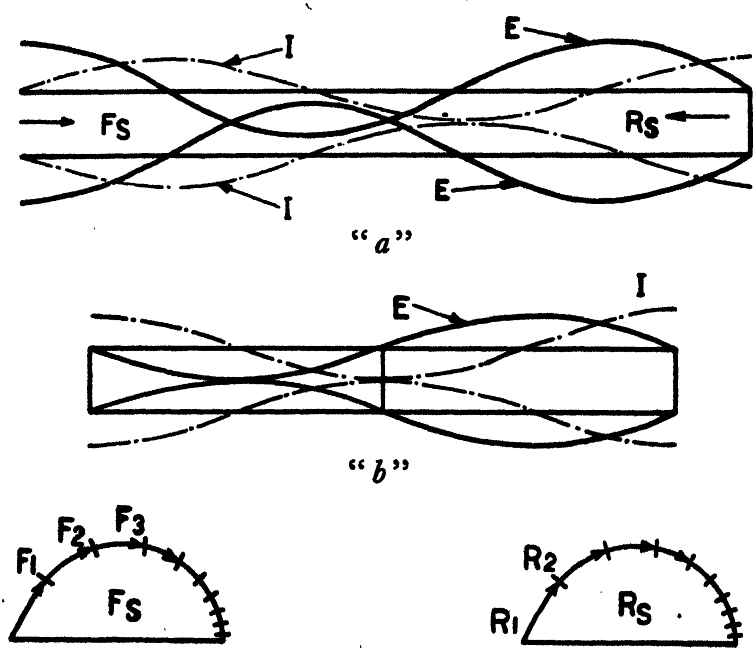


FIGURE 64.

each reflection is in phase with other waves travelling in the same direction, a very large resultant forward wave (and reflected wave) will be formed many times greater in amplitude than the original wave impressed from the generator. In fact in these circumstances it is only the line loss which prevents these resultant waves becoming infinitely great, as can be seen by considering the sum of an infinite number of vectors of equal lengths in line.

If now two waves, both of the same amplitude, each travelling in opposite directions, are combined, they will be found to form a wave stationary in space. In all cases a stationary wave

will consist of a current and voltage wave the nodes of which will be spaced a quarter-wavelength apart in space, the exact positions of these waves on the line being determined by the terminating conditions. In the case of a feeder terminated by a loss-free circuit, there is no transfer of energy along the feeder at all (if its conductor resistance is neglected), but only a surging to and fro of energy. The reader will be familiar with stationary waves as exhibited on strings, etc. If a feeder is terminated by a dissipative circuit, there will be a forward travelling wave supplying the power from the generator. In addition, there may be repeated reflections of energy forming a stationary wave.

The presence of the stationary wave is undesirable because it reduces the amount of energy actually transferred by the feeder for a given generator voltage and also increases the losses because the current in the conductors is increased. We need to know, therefore, how to prevent reflection.

At the far end of the feeder the ratio of voltage to current is fixed by the impedance of the circuit placed there. If then the electric and magnetic components of the advancing wave do not satisfy this relationship, a reflected wave must be formed such that the vector sum of advancing and reflected waves add to give the required condition at the load. The remedy for reflection is, therefore, to provide a circuit whose impedance requires just the relationship between electric and magnetic fields which actually exists in the original advancing wave, and, as will be shown later, this relationship depends upon the line constants. Thus the terminal load should be such that it absorbs the energy at precisely the rate at which it arrives.

Another way of considering the matter is to say that we require an output circuit which, while it may be very different in appearance from the feeder, is from the point of view of the advancing wave the same, so that there is no apparent change of medium and therefore no reflection.

Approximate Equations for Current and Voltage along High Frequency Feeder. When we wish to analyse the behaviour of feeders mathematically it is much more convenient to regard the feeder as made up of distributed inductance and capacity, instead of speaking of electric and magnetic lines in the dielectric. This is only a difference of nomenclature,

of course, and not a physical difference, since when we speak of inductance we mean flux linkages per ampere, and when we speak of capacity we are expressing the electric strain lines per volt. Our feeder can, therefore, be represented by the circuit of Fig. 65, the inductance per unit length being denoted by L and the capacity per unit length by C. Actually the

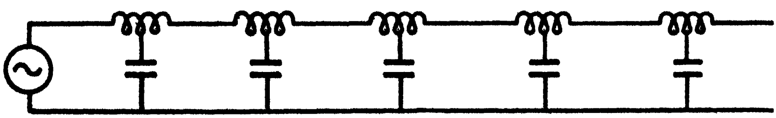


FIGURE 65.

feeder must be more nearly equivalent to the circuit of Fig. 66, because the conductors possess resistance and there must be leakage across insulators. If we applied voice frequencies to our feeder, we should have to take these into account as the telephone engineer has to do, but at high frequencies the inductance and capacity effects are so great compared to the resistance and leakage that the two latter are less important. Thus a simplified approximate analysis, assuming no feeder losses, will be developed to shew the general working of high frequency feeders, complete equations being quoted later without proof when discussing the question of feeder losses.

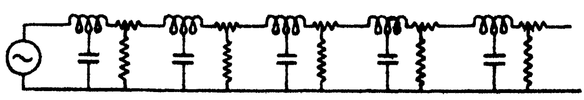


FIGURE 66.

Consider an infinitely long feeder (Fig. 67) supplied from a alternator giving a sinusoidal E.M.F. E_g. Then consider any point at a distance x from the alternator and let the maximum value of the voltage at x be E and maximum value of current at x be I.

Then *decrease* of voltage over δx (assuming current constant over this length)

$$= j\omega LI\delta x$$

$$\text{or } \frac{dE}{dx} = -j\omega LI \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Decrease of current over δx (assuming voltage constant over this length)

$$= j\omega CE\delta x$$

$$\text{or } \frac{dI}{dx} = -j\omega CE \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Differentiating (4) we have

$$\frac{d^2E}{dx^2} = -j\omega L\frac{dI}{dx}$$

$$= (-j\omega L)(-j\omega CE)$$

$$= -\omega^2 LCE \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$\text{Similarly } \frac{d^2I}{dx^2} = (-j\omega C)(-j\omega LI)$$

$$= -\omega^2 LCI \quad . \quad . \quad . \quad . \quad . \quad (7)$$

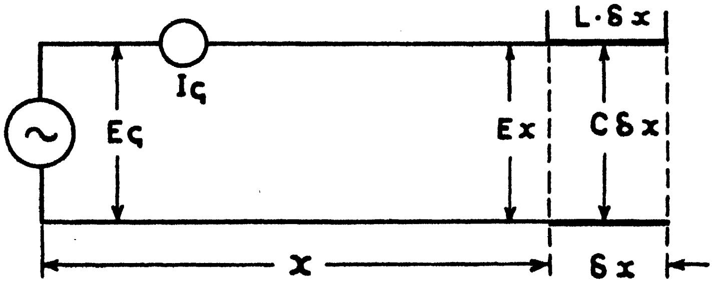


FIGURE 67.

These differential equations are of a well known form, having solutions

$$E = A\cos mx + B\sin mx \quad . \quad . \quad . \quad (8)$$

$$\text{and } I = C\cos mx + D\sin mx \quad . \quad . \quad . \quad (9)$$

where $m = \omega\sqrt{LC}$ and A, B, C and D are constants.

When $x = o, E = E_g$

$$\therefore E_g = (A \times 1) + (B \times o)$$

$$\text{or } A = E_g$$

Differentiating (8)

$$\frac{dE}{dx} = -mE_g\sin mx + Bm\cos mx$$

when $x = o$, $I = I_g$ and $\therefore C = I_g$

Differentiating (9)

$$\frac{dI}{dx} = -I_g\, m \sin mx + Dm \cos mx$$

But $$\frac{dE}{dx} = -j\omega LI$$

and hence

$$-mE_g \sin mx + Bm \cos mx = -j\omega L\,(I_g \cos mx + D \sin mx)$$

As this is true for any value of x it follows that sine and cosine terms must be equal.

$$-mE_g = -j\omega LD$$

$$D = \frac{mE_g}{j\omega L} = \frac{\omega\sqrt{LC}\,E_g}{j\omega L} = -jE_g\sqrt{\frac{C}{L}}$$

and $mB = -j\omega LI_g$

$$B = -\frac{j\omega LI_g}{m} = -\frac{j\omega LI_g}{\omega\sqrt{LC}}$$

$$= -j\sqrt{\frac{L}{C}}\; I_g$$

Hence equations become

$$E_s = E_g \cos mx - j\sqrt{\frac{L}{C}}\, I_g \sin mx \quad . \quad . \quad . \quad . \quad (10)$$

$$I_s = I_g \cos mx - j\sqrt{\frac{C}{L}}\; E_g \sin mx \quad . \quad . \quad . \quad . \quad (11)$$

We cannot eliminate the unknown I_g unless we know the terminating conditions of the line.

Suppose the line to be infinitely long, then no reflection is possible. Let Z_0 then be the effective impedance of the feeder from the generator end. Then $I_g = \frac{E_g}{Z_0}$. The impedance of the feeder at any finite distance x from the generator must still be Z_0, because there is still an infinite length of the feeder lying beyond the point considered. In order to find Z_0 take a value of x such that $mx = \pi/2$, so that $\cos mx = o$ and $\sin mx = 1$.

$$\text{Then}\quad \frac{E_{x1}}{I_{x1}} = Z_0 = \frac{-j\sqrt{\frac{L}{C}}\cdot\frac{E_g}{Z_0}}{-j\sqrt{\frac{C}{L}}\;E_g}\,.$$

$$\text{or}\quad Z_0^2 = \frac{L}{C}$$

$$Z_0 = \sqrt{\frac{L}{C}}\quad . \quad . \quad . \quad . \quad . \qquad (12)$$

Hence the infinite line loads the generator in exactly the same manner as a pure resistance of value $\sqrt{L/C}$. This value may be denoted by R_0 and is termed the characteristic resistance of the feeder.

It is important to notice that although the value of R_0 in ohms is given by $\sqrt{L/C}$ (where L and C are expressed in henries and farads respectively), yet R_0 is an effective resistance, not a reactance. We have also shown that the impedance at any point along the infinite feeder is R_0 and therefore the current is in phase with the voltage and of value $\frac{E_0}{R_0}$. Suppose now we have a finite length of feeder, then if we place a resistance equal to R_0 across the far end there will be no reflection because we have placed there a resistance equivalent in every way to the impedance of the line to infinity. From another point of view we have inserted a termination which behaves in the same way to the wave as the feeder behaves, and therefore there is no reflection because there is no change of medium. Evidently this is the correct termination to use for most efficient transmission of energy, since no energy will now be wasted in a system of stationary waves superimposed on the travelling wave. Notice also that since the terminating resistance R_0 is replacing the portion of feeder cut off, which extends always to infinity, the termination is correct anywhere; and hence the actual length of feeder used is of no account and the correct termination will always be R_0. Because the feeder has actually resistance and insulation loss, the correct termination will be a complex impedance, to determine which the full analysis of the transmission line as used by the telephone

engineer would have to be applied, but the difference in the case of short wave feeders is small.

The equations for the voltage and current at any point in an infinite or correctly terminated feeder may, therefore, be written

$$E_x = E_g\,(\cos \omega\sqrt{LC}\,x - j \sin \omega\sqrt{LC}.x) \qquad . \qquad . \quad (13)$$

$$I_x = \frac{E_g}{R_0}\,(\cos \omega\sqrt{LC}\,x - j \sin \omega\sqrt{LC}.x) \qquad . \qquad . \quad (14)$$

These expressions will evidently repeat themselves every time $\omega\sqrt{LC}.x$ is increased by 2π or when x is increased by $\dfrac{2\pi}{\omega\sqrt{LC}}$. In other words the wavelength $\lambda = \dfrac{2\pi}{\omega\sqrt{LC}}$ (15)

For any wave motion $v = f\lambda$

$$\therefore v = \frac{\omega}{2\pi} \cdot \frac{2\pi}{\omega\sqrt{LC}} = \frac{1}{\sqrt{LC}} \quad . \quad . \quad . \quad . \quad (16)$$

This velocity is equal to that of a "free" electro-magnetic wave in a pure dielectric, the conductors having been assumed to be mere guides in which no loss is taking place.

The high frequency feeder is what a telephone engineer terms a distortionless line, since it behaves the same to all frequencies sufficiently high that the assumptions made are reasonably true. There is no attenuation of any frequency within this range and the velocity is the same for this range of frequencies.

We have thus far considered the feeder terminated with an actual resistance but except for test purposes such a case would not be met with in practice. High frequency feeders normally will be terminated by loaded resonant circuits of some form and it is well known that circuits of this type can act as a resistive load.

By the design of suitable coupling circuits, which will be described later, such terminations may be made to match the feeder characteristic resistance exactly.

An interesting relationship exists *in a loss-free line* between the voltage at any point and the current at a point one quarter-wavelength away (in either direction).

If x is increased by $\dfrac{\lambda}{4}$, then mx is increased by $\dfrac{\pi}{2}$, so that

$$E_x = E_g \cos mx - j\, R_0\, I_g \sin mx$$

and $$I_{x+\frac{\lambda}{4}} = I_g \cos\left(mx + \frac{\pi}{2}\right) - j\,\frac{E_g}{R_0} \sin\left(mx + \frac{\pi}{2}\right)$$

$$= -I_g \sin mx - j\,\frac{E_g}{R_0} \cos mx$$

$$= -j\, E_x/R_0 \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Hence, whatever may be the terminating conditions in a loss-free line, the voltage at any point is equal (in magnitude) to the current at a point $\frac{\lambda}{4}$ away, multiplied by the characteristic resistance. This relationship has been used to estimate high voltages on such a line by means of current measurements, which are easier to make at high frequencies.

Value of Characteristic Impedance R_0. A number of methods are available for a direct measurement of the characteristic impedance of a line, but since the value of R_0 depends almost entirely upon the ratio of the inductance and capacity per unit length of feeder we can calculate its value accurately and in Appendix III curves are shown which give the value of R_0 for twin and four-wire open lines, and concentric-tube lines, from which we can observe that twin lines have an R_0 of the order of 600 ohms, four-wire lines 350 ohms, and concentric lines 80 to 100 ohms, as usually designed.

We have previously mentioned the stationary wave phenomenon with lines not correctly terminated and we will now study the effects produced when open or shorted lines of finite length are connected across an alternator.

Open Circuited Feeder. Suppose now that we have a feeder of length x_1, which has its end open circuited instead of being connected across a resistance of value R_0. Then voltage and current equations at x_1 become

$$E_{x1} = E_g \cos mx_1 - j\, R_0\, I_g \sin mx_1 \quad . \quad . \quad . \quad (18)$$

$$0 = I_g \cos mx_1 - \frac{j}{R_0} E_g \sin mx_1 \quad . \quad . \quad . \quad . \quad (19)$$

From (19) $I_g \cos mx_1 = j\,\frac{E_g}{R_0} \sin mx_1$

$$I_g = j\,\frac{E_g}{R}\tan mx_1$$

The impedance of the feeder from the generator end is therefore

$$-j\, R_o \cot mx_1 \qquad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

$$\text{or } -j \sqrt{\frac{L}{C}} \cot \omega \sqrt{LC}\, x_1 \qquad . \quad . \quad . \quad . \quad . \quad (21)$$

Hence when x_1 is of such a value as to make $\cot \omega \sqrt{LC}.x_1$ positive, the open-circuited feeder loads the generator as a capacity reactance of value $\sqrt{\frac{L}{C}} \cot \omega \sqrt{LC}\, x_1$, while when the co-tangent becomes negative the reactance becomes inductive. Using the value of λ obtained in (15), this may be re-written

$$X = -R_o \cot \frac{2\pi}{\lambda} x_1 \qquad . \quad . \quad . \quad . \quad . \quad (22)$$

substituting the value obtained for I_g in the general equations, the equations for voltage and current at any point x become :

$$E_x = E_g (\cos mx + \tan mx_1 \sin mx) \qquad . \quad . \quad . \quad (23)$$

$$I_x = \frac{E_g}{R_o} (\tan mx_1 \cos mx - \sin mx) \qquad . \quad . \quad . \quad (24)$$

Since $\tan mx_1$ is a constant, there will be a series of values of x which make the equations zero. These values will occur when mx is increased by π, that is when the line length is increased by $\frac{\lambda}{2}$.

Thus along an open-circuited line of many wavelengths there will be a series of points $\frac{\lambda}{2}$ apart, where the voltage is always zero. These are termed voltage nodes. There will be a similar set of points where the current is zero, and these will occur midway between the voltage nodes. For any line length, however, there must be a current node and a voltage anti-node at the open end, and the positions of corresponding nodes and anti-nodes can be reckoned back from the end. The magnitude of these stationary waves however and the relative current and voltage values from the generator will depend upon the length of line (and line attenuation). If there was really no loss in the feeder the wave produced will be entirely of the stationary type, there being no transmission of energy along the feeder at all and such a system

of waves could be sustained without the expenditure of any energy.

It will be observed that because of these stationary waves there is a spatial distribution of line current and voltage of one-quarter wave-length, and a quadrature time phase between them. In consequence the ratio $\frac{E}{I}$ or effective impedance at different points along the line will vary very greatly in magnitude. For instance, at the open end, where I is zero, the

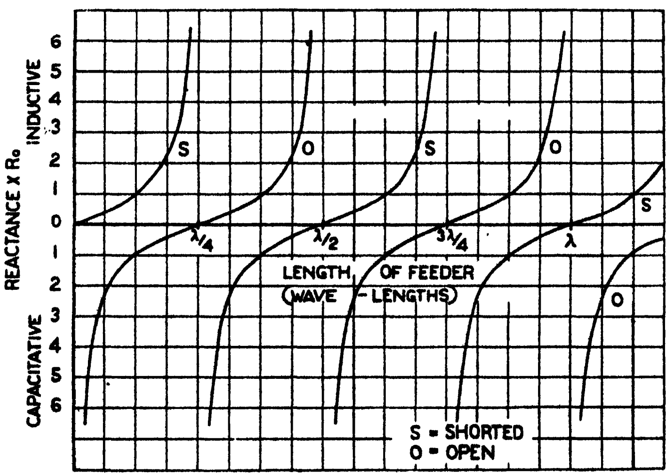


FIGURE 68.

effective impedance is infinitely great and resistive in character since I is zero, whereas at a point one-quarter wave back from the end the reverse holds, namely, E is now zero and the ratio $\frac{E}{I}$ indicates a resistive load of zero impedance. The values of $\frac{E}{I}$ at points between these extremes show perfect reactive conditions to exist, the sign of reactance changing abruptly as one passes through each quarter wave position.

Fig. 68 shows these variations of reactance with feeder length from which it may be seen that an open circuited feeder one-quarter wavelength long (or any odd multiple

thereof) throws back across the generator a short circuit; A length of half a wavelength (and multiples thereof) acts as an open circuit; and lengths of one-eighth wavelength (and odd multiples thereof) a pure reactance equal in value to R_o, but changing sign every quarter wavelength. It is often convenient when setting up feeder and aerial systems to use a length of feeder as a matching reactance because its value can be accurately predetermined by length and it avoids the use of an ordinary condensers and inductances in the open.

Short Circuited Feeder. If the end of a feeder of length x_1 be closed by a "loss free" conductor, then at x_1 equations become

$$0 = E_g \cos mx_1 - j R_o I_g \sin mx_1 \qquad . \qquad . \qquad (25)$$

$$\text{and } I_{x1} = I_g \cos mx_1 - j \frac{E_g}{R_o} \sin mx_1 \quad . \quad . \quad . \quad (26)$$

$$I = \frac{E_g \cos mx_1}{j R_o \sin mx_1}$$

$$= -j \frac{E_g}{R_o} \cot mx_1$$

The reactance of the feeder, considered from the generator end, will therefore be

$$j R_o \tan mx_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

Hence when tan mx_1 is positive the feeder will behave as an inductance, the current lagging 90°, but when the tangent becomes negative the feeder will be equivalent to a capacity. Substituting for I_g in the current equation

$$I_{x1} = -j \frac{E_g}{R_o} \left\{ \cot mx_1 \cos mx_1 + \sin mx_1 \right\} \qquad (28)$$

The equation for the voltage at any point x becomes

$$E_x = E_g \cos mx - j R_o \left(\frac{-jE_g}{R_o} \cot mx_1 \right) \sin mx$$

$$= E_g (\cos mx - \cot mx_1 \sin mx)$$

The current equation becomes

$$I_x = -j \frac{E_g}{R_o} (\cot mx_1 \cos mx + \sin mx) \quad . \quad . \quad (29)$$

Evidently when $\cos mx = \cot mx_1 \sin mx$, the expression for E_x will become zero, indicating that the maximum value of E for

this value of x is zero, i.e. the voltage at this point is always zero.

As cot mx_1 is a constant these nodes will occur every time mx is increased by π, i.e. at distances separated by $\frac{\lambda}{2}$. The actual distances from the generator end at which the nodes come will depend upon x_1, the length of the feeder.

The way in which the reactance varies with length is shown in Fig. 68, from which it is evident that there is but little difference between the behaviour of an open-circuited and a shorted line, except that the stationary wave system is moved forward one-quarter wavelength.

Feeder Terminated by any Impedance Z_r. No Line Attenuation.

If the feeder is of length l then

$$E_1 = E_g \cos ml - j\, R_o I_g \sin ml \qquad \text{(see 10, 11, and 12)}$$

$$I_1 = I_g \cos ml - j\, \frac{E_g}{R_o} \sin ml$$

But if the feeder is terminated by an impedance Z_r, then $E_1 = I_1 Z_r$ and $E_g \cos ml - j\, R_o I_g \sin ml = I_g Z_r \cos ml - j\, E_g \frac{Z_r}{R_o} \sin ml$

from which $I_g = E_g \dfrac{\left(\cos ml + j \dfrac{Z_r}{R_o} \sin ml\right)}{Z_r \cos ml + j\, R_o \sin ml}$

Evidently by substituting this value for I_g in the general equations (10) and (11) we can obtain expressions for the voltage and current at any point along a line terminated by an impedance Z_r. We are usually more concerned to know what is the "sending-end" impedance of such a line, that is, the impedance which it places across the generator.

This is evidently given by

$$\frac{E_g}{I_g} \text{ that is, } Z = \frac{Z_r \cos ml + j\, R_o \sin ml}{\cos ml + j \dfrac{Z_r}{R_o} \sin ml} \qquad . \quad . \quad (30)$$

$$= R_o \frac{Z_r + j\, R_o \tan ml}{R_o + j\, Z_r \tan ml} \qquad . \quad . \quad (31)$$

Properties of a Quarter-Wave Line. We have already seen that a $\frac{\lambda}{4}$ line, open-circuited at the far end, presents zero impedance at the generator end. A $\frac{\lambda}{4}$ line has a number of other properties which are useful and interesting.

For instance, if a $\frac{\lambda}{4}$ line is terminated by a pure resistance $\frac{R_o}{n}$, Fig. 69, then the input impedance is a resistance of value $n\,R_o$. This may be proved as follows:

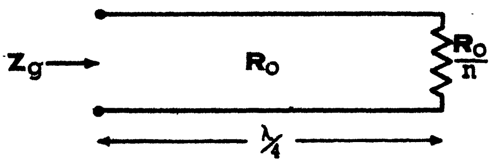


FIGURE 69.

Re-writing (31) in terms of "electrical length"

$$Z_g = R_o \frac{Z_r + j\,R_o \tan\theta}{R_o + j\,Z_r \tan\theta}$$

which may be re-arranged as

$$Z_g = R_o \frac{Z_r \cos\theta + j\,R_o \sin\theta}{R_o \cos\theta + j\,Z_r \sin\theta} \quad . \quad . \quad . \quad . \quad (32)$$

For a $\frac{\lambda}{4}$ line $\left(\theta = \frac{\pi}{2}\right)$ this becomes

$$Z_g = \frac{R_o^2}{Z_r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

Hence if Z_r is $\frac{1}{n}R_o$ $\qquad Z_g = n\,R_o$ . . . (34)

A similar argument will show that a terminating inductive reactance of value $\frac{R_o}{n}$ appears at the generator end of a $\frac{\lambda}{4}$ line as a capacitative reactance of value $n\,R_o$, and the converse.

A $\frac{\lambda}{4}$ line can therefore be used as a transformer for matching low impedances to high values or vice versa. A common

application of such a transformer is to terminate a feeder of characteristic resistance R_{o1}, supplying an aerial, equivalent to a resistance R, as shown in Fig. 70. A $\frac{\lambda}{4}$ length of feeder, R_{o2}, having the correct characteristic resistance will terminate the main feeder correctly.

There will be stationary waves set up in the $\frac{\lambda}{4}$ line because it is not correctly terminated, but the conductor and insulator losses thereby introduced are not serious because of the short length concerned.

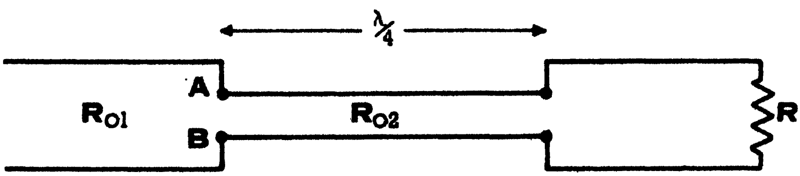


FIGURE 70.

From (33),

$$Z_{AB} = \frac{R_{o2}^2}{R} \text{ and } Z_{AB} \text{ is to equal } R_{o1}$$

$$\therefore\ R_{o2} = \sqrt{R\,R_{o1}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

It is also possible to transform a complex impedance into a pure resistance by altering the length of line somewhat. Hence, for example, an aerial which is equivalent, at the feed point, to an impedance $R + j\,X$ can be arranged to give a pure resistance R_o, at AB.

From (31)

$$Z_{AB} = R_{o2}\,\frac{(R + j\,X + R_{o2}\tan\theta_2)}{R_{o2} + j\,(R + j\,X)\tan\theta_2}$$

Rationalising and re-arranging:

$$Z_{AB} =$$

$$R_{o2}\,\frac{RR_{o2}(1+\tan^2\theta_2)+j\{R_{o2}X(1-\tan^2\theta_2)+\tan\theta_2\,(R_{o2}^2-X^2-R^2)\}}{(R_{o2}-X\tan\theta_2)^2+R^2\tan^2\theta_2} \quad . \quad . \quad (35)$$

Since R_{o1} is a pure resistance and we are trying to make $Z_{AB}=R_{o1}$, it follows that the "imaginary" part of this expression should equal zero, and this fact will give us one relationship for the design of the transforming length of feeder.

$$R_{o2}X\,(1-\tan^2\theta_2)+\tan\theta\,(R_{o2}^2-X^2-R^2)=0$$

Multiplying by $\cos^2\theta_2$ and re-arranging :

$$\sin\theta_2\cos\theta_2\,(R_{o2}^2-X^2-R^2)=-R_{o2}X\,(\cos^2\theta_2-\sin^2\theta_2)$$

$$\tfrac{1}{2}\,(R_{o2}^2-X^2-R^2)\sin 2\theta_2=-R_{o2}X\cos 2\theta_2$$

$$\tan 2\theta_2=-\frac{2\,R_{o2}\,X}{R_{o2}^2-X^2-R^2}\quad . \quad . \quad . \quad . \quad (36)$$

This expression for the length of the transforming feeder contains its characteristic impedance and it will therefore be necessary to assume a value for R_{o2} and solve for θ_2. The value found can then be inserted in the equation

$$Z_{AB}=\frac{R\,R_{o2}^2\,(1+\tan^2\theta_2)}{(R_{o2}-X\tan\theta_2)^2+R^2\tan^2\theta_2}\quad . \quad . \quad (37)$$

to see how nearly $Z_{AB}=R_{o1}$, the condition we wish to achieve. If the difference is too great, another attempt must be made with a different value of R_{o2}.

It is possible, however, to obtain an expression for R_{o2} which does not involve θ_2, and this can be shown to be

$$R_{o2}^2=R_o\,R\left\{1-\frac{X^2}{R\,(R_o-R)}\right\}\quad . \quad . \quad . \quad (38)$$

Example 1. A half-wave aerial, tuned to 20 megacycles, is equivalent (between its feedpoints), to a non-inductive resistance of 100 ohms. It is to be fed by a parallel-wire feeder having a characteristic impedance of 500 ohms. Design a suitable transforming feeder to insert between the main feeder and the aerial so that the former may be correctly terminated.

The feeder should be $\frac{\lambda}{4}$ long. If the velocity along it may be taken to be the same as in free space, then the required length is 3.75 metres. (Actually, the velocity will be somewhat less due to losses and to the presence of supporting insulators, and the correct length will therefore be reduced. A usual value of velocity for parallel-wire feeders supported at intervals by good insulators is some 90% of that in free space.)

The characteristic impedance of the transforming feeder should be $\sqrt{500 \times 100} = 224$ ohms. Reference to Appendix III shows that this could be realised by using two rods of 0.5 cm. radius, spaced 3.2 cms. apart.

Example 2. Conditions as in example (1) except that impedance of aerial at feed points is equivalent to a resistance of 100 ohms in series with an inductive reactance of 50 ohms.

$$\text{From (38), } R_{o_2}^2 = 500 \times 100 \left\{ 1 - \frac{50^2}{100\,(500 - 50)} \right\}$$

$$R_{o_2} = 218 \text{ ohms.}$$

$$\text{From (36), } \tan 2\theta = -\frac{2 \times 50 \times 218}{218^2 - 100^2 - 50^2}$$

$$= -0.628$$

$$2\theta = 148^\circ \qquad \theta = 74^\circ$$

A general theorem may here be stated, that is frequently useful in connection with feeder and aerial calculations. Any impedance can evidently be expressed either as a resistance and a reactance in series or a different resistance and reactance in parallel. Thus consider the circuits (*a*) and (*b*) of Fig. 71.

If (*a*) represents an actual arrangement of condenser and resistance in series it is nevertheless possible to find values of a condenser and resistance which placed in parallel would take the same current as the series circuit when the same voltage at the same frequency was applied to it.

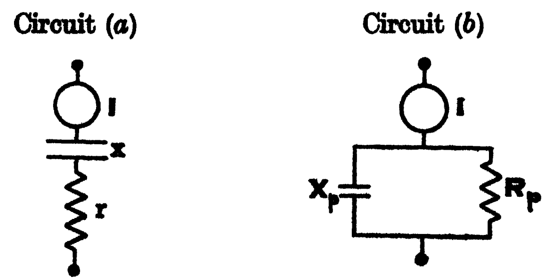


FIGURE 71.

$$I = \frac{E}{r - jx} = \frac{E\,(r + jx)}{r^2 + x^2} \qquad I = \frac{E}{R_p} + \frac{E}{-jX_p}$$

$$\text{Hence } \frac{r}{r^2 + x^2} = \frac{1}{R_p} \text{ and } \frac{x}{r^2 + x^2} = \frac{1}{X_p}$$

or $R_p = \dfrac{r^2 + x^2}{r}$ and $X_p = \dfrac{r^2 + x^2}{x}$ . . . (39)

Conversely the conversion of parallel elements of X_p and R_p into equivalent series elements r and x, can be carried out by the following formulæ:—

$$r = R_p \frac{X_p^2}{R_p^2 + X_p^2} \quad . \quad . \quad . \quad . \quad . \quad (40)$$

$$x = X_p \frac{R_p^2}{R_p^2 + X_p^2} \quad . \quad . \quad . \quad . \quad . \quad (41)$$

or if r has previously been found x can be obtained more simply from

$$x = \frac{R_p r}{X_p} \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

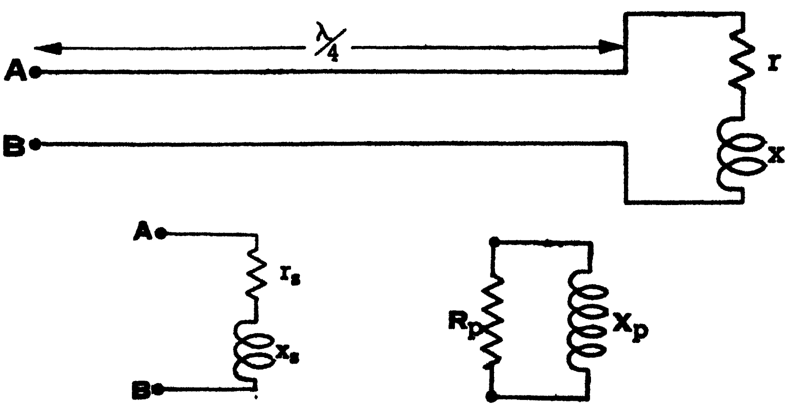


FIGURE 72.

By the use of this theorem, together with our knowledge of the properties of a $\frac{\lambda}{4}$ line, we can deduce the values of r and x by measuring the impedance between A and B as a parallel combination R_p and X_p, Fig. 72.

From (33) $r_s + jx_s = \dfrac{R_o^2}{r + jx}$

from which $r_s = \dfrac{R_o^2 r}{r^2 + x^2}$ and $x_s = \dfrac{-R_o^2 x}{r^2 + x^2}$

From (39) $R_p = \dfrac{r_s^2+x_s^2}{r_s} = \dfrac{R_o^2}{r}$

$$X_p = \frac{r_s^2+x_s^2}{x_s} = \frac{-R_o^2}{x}$$

$$\text{or } r = \frac{R_o^2}{R_p} \text{ and } x = \frac{-R_o^2}{X_p} \qquad . \quad . \quad . \qquad (43)$$

This relationship is sometimes useful when conducting measurements on feeders.

The Reactance "Transformer." An alternative method for transforming an impedance to another value involves the use of what is known as a reactance transformer. For

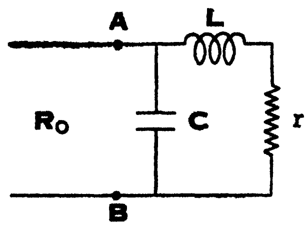


FIGURE 73.

instance, suppose we have a feeder having a characteristic resistance R_o, Fig. 73, and desire to couple to it a load resistance r whose value is but $\dfrac{1}{n} R_o$.

By placing an inductance in series with r we can increase the impedance of this branch but there will then be a large reactive component. This can be balanced by the parallel capacity leaving the net impedance between A and B purely resistive and of the correct value.

The correct sizes of X_L and X_C can be determined as follows:

By the usual rule for parallel impedances,

$$R_{AB} = nr = \frac{(r + j\,X_L)\ \ (-j\,X_C)}{r + j\,X_L - jX_C}$$

$$nr^2 + jnr\,(X_L - X_C) = X_L X_C - jX_C\,r$$

equating the real parts, $nr^2 = X_L X_C$ or $X_L = \dfrac{nr^2}{X_C}$

equating the unreal parts, $nr\,(X_L - X_C) = -X_C r$

$$n\frac{nr^2}{X_C} - nX_C = -X_C$$

$$X_C = \frac{nr}{\sqrt{n-1}} \quad . \quad . \quad . \quad (44)$$

$$\text{and} \quad X_L = r\sqrt{n-1} \quad . \quad . \quad . \quad (45)$$

It is not actually necessary that the load impedance should be simply a pure resistance r. If the load has actually inductive reactance for example, then X_L may be reduced to give the same total inductive reactance for the branch.

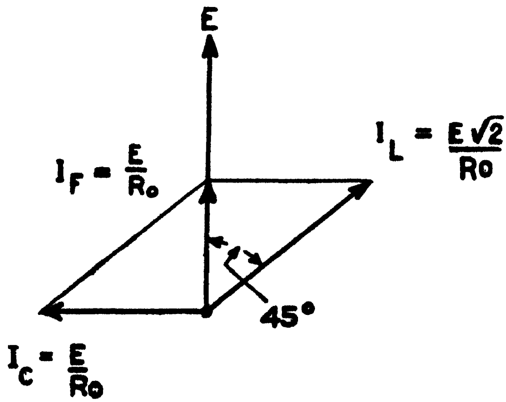


FIGURE 74.

The arrangement can, of course, be used in the reverse direction. Thus the larger resistance, R, may be the load which is to be transformed into a smaller resistance, r.

Examples of the Use of the Reactance Transformer. (1) A main feeder branches into two, all the feeders being of the same concentric type with a characteristic impedance of R_o. It is desired to install a reactance transformer at the junction so that the main feeder may be correctly terminated.

The load impedance of the two feeders in parallel will be $\frac{R_o}{2}$. Hence $n = 2$ and from our equations we find that $X_C = R_o$ and $X_L = \frac{R_o}{2}$. The vector diagram for this simple case is given in Fig. 74.

(2). The base impedance of a given aerial may be considered as a resistance of 300 ohms in series with a capacity reactance of 200 ohms. The feeder used has a characteristic impedance of 75 ohms. Determine the reactances to be used in a suitable transformer.

This is evidently a transformation in the reverse direction to that discussed previously and therefore the connections will be as Fig. 75. It will be necessary to express the aerial impedance in parallel form and from equations (39) these become 433 ohms resistance and 650 ohms capacity reactance. The reactance will

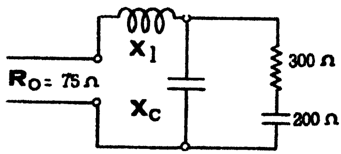


FIGURE 75.

be provided for by increasing the size of X_C and will therefore be ignored at the moment.

$$n = \frac{433}{75} = 5.78$$

$$X_C = \frac{433}{\sqrt{4.78}} = 198 \text{ ohms and } X_l = 75\sqrt{4.78} = 164 \text{ ohms.}$$

It will be seen that X_C is made up of two parallel reactances, 650 ohms due to the aerial and X, say, due to the condenser in the transformer. We therefore have $\frac{1}{650} + \frac{1}{X} = \frac{1}{198}$ from which $X = 286$ ohms.

Losses in Feeders. Any actual feeder must, of course, attenuate the energy passing along it because of conductor and insulator losses and energy may also be lost by radiation.

Conductor loss may be calculated with fair accuracy but insulator losses are not amenable to calculations. With the parallel-wire feeder there may be considerable radiation loss of amount depending upon a number of factors.

Suppose a power P is to be delivered by a feeder of characteristic impedance R_o and conductor resistance r. Then the feeder current $I = \sqrt{P/R_o}$ and voltage $V = \sqrt{P R_o}$. Conductor losses are $P\frac{r}{R_o}$ and hence, if these only are considered, equal transmission efficiencies will be obtained from different types of feeder having the same value for r/R_o.

The feeder having the larger R_0 will tend to have greater dielectric losses because the voltage is higher.

It has already been mentioned that concentric-tube feeders have a characteristic resistance of between 80 and 100 ohms, and parallel-wire feeders some 600 ohms, and 4-wire feeders some 350 ohms, so that the former carry larger currents for a given power and therefore need to have greater conductor section. In both cases an increase of spacing of the conductors increases the characteristic resistance, but in the parallel wire case, if we increase the spacing too much it becomes more difficult to prevent radiation. With concentric-tube feeders any diameter of outer tube increases materially the cost of the feeder, whilst there is an optimum ratio of inner and outer diameters giving best efficiency for a given outer tube diameter.

That there will be an optimum ratio is evident from the following considerations. Since the outer diameter is fixed, the conductor loss in the outer tube is fixed. The larger the radius of inner tube the larger is its conducting surface, and therefore the smaller the conductor resistance. By increasing the inner tube diameter, however, we bring the two tubes closer together and thereby decrease the characteristic resistance, which means more current for a given power. If we increase the inner tube diameter beyond a certain value, therefore, the reduction of conductor resistance is more than offset by the decrease of characteristic resistance. This optimum ratio is determined in Appendix IV and found to be 3·6.

Copper Losses in Feeders. In Appendix IV it is also shown that the power efficiency of a km. length of a concentric tube feeder (taking account of copper losses only) is given by

$$\log_{10} \eta = -1{\cdot}30 \times 10^{-5} \sqrt{f}\, \frac{\left(\dfrac{1}{r_1} + \dfrac{1}{r_2}\right)}{\log_{10} \dfrac{r_2}{r_1}} \quad . \quad (46)$$

whilst the loss per km. in dbs. is

$$1{\cdot}30 \times 10^{-4} \sqrt{f}\, \frac{\left(\dfrac{1}{r_1} + \dfrac{1}{r_2}\right)}{\log_{10} \dfrac{r_2}{r_1}} \quad . \quad . \quad . \quad (47)$$

where η =efficiency.

f =frequency in cycles per second.

r_1 =radius of inner tube in cms.

r_2= ,, outer ,, ,,

The efficiencies per km. length and loss in dbs. per km. of some typical concentric tube feeders, calculated from the above formulæ, are given in Table I:

TABLE I—FREQUENCY = 20 MEGACYCLES.

Type.	Outer Radius of Inner Tube (Cms.)	Inner Radius of Outer Tube (Cms.)	Ratio $\frac{r_2}{r_1}$	Percentage Efficiency per Km.	Attenuation Dbs. per Km.	R_0
No. 0	3·30	13·0	4	81·3	·95	83
,, 1	1·11	4·44	4	77·8	1·09	83
,, 2	0·875	3·17	3·6	70·5	1·52	77·0
,, 3	0·238	0·795	3·34	24·8	6·05	72·2

Corresponding calculations for some arrangements of parallel wire feeders are given in Table II. The calculation neglects losses due to earth currents and also the effect the presence of the other wire has upon the current distribution.

TABLE II—FREQUENCY = 20 MEGACYCLES.

Wire Gauge No.	Radius of Wire (Cms.)	Distance between Wires (Cms.)	Ratio d/r	Percentage Efficiency per Km.	Attenuation Dbs. per Km.	R_0
		10	41·1	71·1	1·48	445
6	0·243	20	82·2	75·0	1·25	530
		30	123	76·9	1·14	578
	—	10	49·3	67·8	1·69	468
8	0·203	20	98·6	72·8	1·43	551
	—	30	148	73·8	1·32	600
	—	10	75·7	58·3	2·34	519
12	0·132	20	151	63·0	2·01	602
		30	227	65·2	1·86	650

The attenuation at any other frequency f_1 may be found (as equation shows) by applying the ratio $\sqrt{\frac{f_1}{20\times10^6}}$ to the results given for 20 megacycles.

Feeder Attenuation Measurements. The results of measurements on concentric tube feeders indicates that with large feeders the total loss is of the order of double the copper

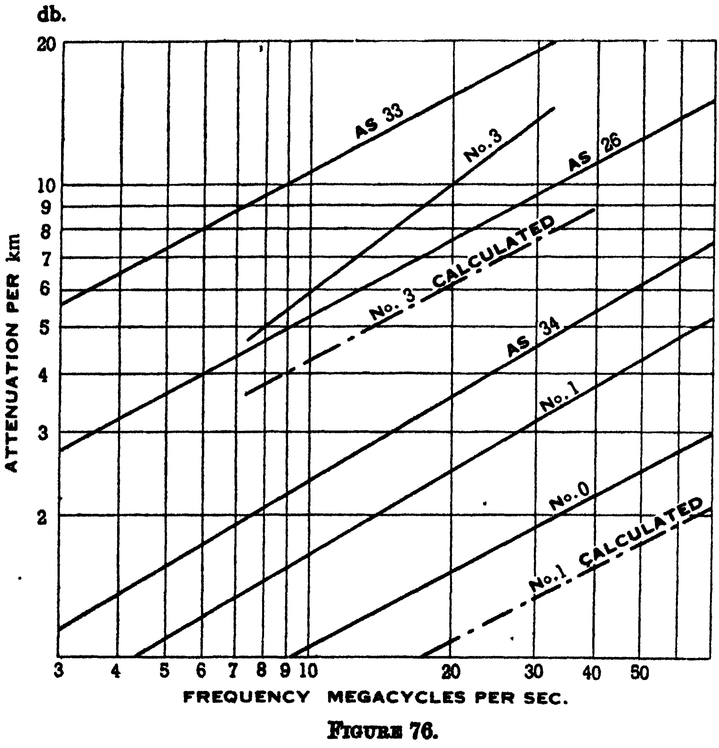


FIGURE 76.

loss, whereas with small feeders the total loss is less than double owing to the greater proportion of copper loss with the small diameter inner conductor used. In Fig. 76 are shown curves for various types of feeder, and in two cases, curves for both calculated and measured loss are shown, as indicated in the Figure.

Effect of Insulator Spacing. All our calculations assume a uniformly distributed capacity between the lines but solid

insulation must be employed to keep the conductors in position and these insulators will, in most cases, be spaced at intervals in order to reduce dielectric loss and cost. Since all known materials have a dielectric constant greater than that of air, we are therefore inserting additional parallel capacities at intervals, which may have values up to about $5\mu\mu F$.

The effect of the insulators is very dependent upon their spacing in relation to the wavelength being used.

Consider a feeder (Fig. 77) in which the insulators are spaced $\frac{\lambda}{2}$ apart. It can be seen from (31) that if $ml = \pi$, then $Z_g = Z_r$. That is, a $\frac{\lambda}{2}$ line does not transform impedances but a terminating Z_r produces an impedance Z_r across the generator. Hence C_3 is effectively in parallel with C_2 to give

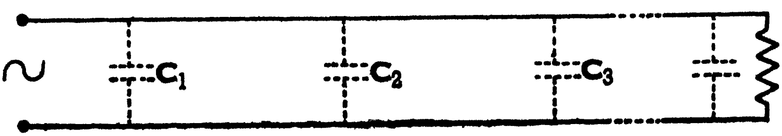


FIGURE 77.

$2C$ and this will produce the equivalent of $3C$ at C_1, so that if there are n insulators along the feeder, a capacity of nC is placed across the generator, and the input impedance is very different from R_0 even if the feeder is correctly terminated, and will depend greatly upon the length of the feeder. There will be large reflected waves set up at the nearer insulators where low capacity reactances are effectively shunted across the feeder. Fig. 78 shows the variation of input impedance looking towards the terminal load as the distance along the feeder is increased. If insulators are spaced at $\frac{\lambda}{4}$ intervals, the capacity reactance of C_1 becomes an inductive reactive at C_2, partially cancelling the effect of C_2, and so on. No cumulative effects occur and the variation of input impedance with length is therefore small. Such an arrangement would be, however, critical with frequency. It is more usual therefore to increase the number of insulators considerably, whereby the change of input impedance with length can be reduced

to a negligible amount, eleven per wavelength being a usual figure. This question of insulator spacing principally concerns ultra-short wave feeders because at the longer wavelengths such spacings as $\frac{\lambda}{2}$ or $\frac{\lambda}{4}$ would be quite inadequate mechanically.

Velocity Slip. Another and important feature of a feeder which is controlled by the dielectric material between the feeder line is the reduction of velocity of the wave in the feeder relative to the wave in free space. For instance, we may require to use a quarter wavelength piece of feeder and it is

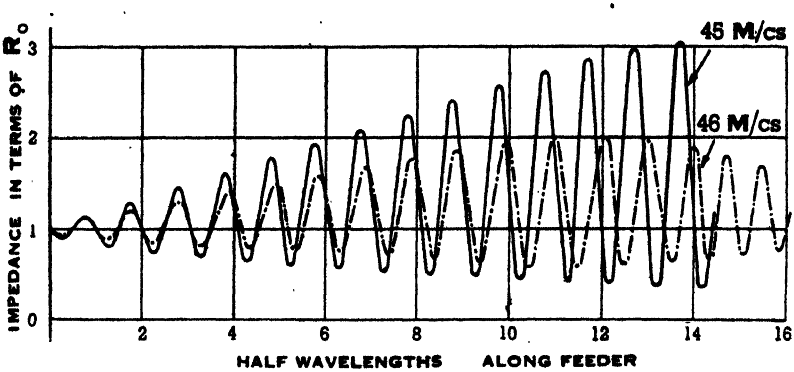


FIGURE 78.

essential to know the actual length of feeder required to act as a quarter wavelength. In all cases the actual length will be shorter by a percentage which is dependent upon the reduction of velocity of the wave in the feeder and may be termed the "velocity-slip." Thus feeders having a continuous dielectric will have greater velocity slip than feeders having air spacing with occasional insulator supports, but it should be remarked that a large amount of velocity slip does not necessarily mean a proportional insulator loss because it is possible to produce insulating materials having a high dielectric constant and yet having low loss. Values for velocity slip in typical feeders are shown in Table III.

Types of Concentric Feeder. Concentric feeders used at transmitting stations are generally constructed in rigid form

TABLE III

Maker	Type	Diameter Outer	Diameter Inner	Insulation	R_o	Capacity per metre $\mu\mu$F	Attenuation α db. per km.	Velocity Slip
T.C.M.	AS 34	3″	0·5″	Steatite Discs	92	41·6	$\cdot 549 f^{\cdot 595}$	11·5%
	AS 33	5″	0·64″	Trolitul Discs	100	34·0	$3{\cdot}12 f^{\cdot 23}$	4%
	AS 26	1·0″	·127″	Trolitul Discs	100	32·5	$\cdot 56 f^{\cdot 53}$	2·5%
	PT 1	·188″	·022″	Solid "Telcothene"	78	68·9	$10 f^{\cdot 69}$	30%
	Peroflex	·156″	23 S.W.G.	"Peroflex" Tube	70	90·0	$15{\cdot}71 f^{\cdot 7}$	26·6%
	Trolitul String	·156″	23 S.W.G.	Trolitul String	76	45·0	$8 f^{\cdot 6}$	12·6%
SIEMENS	E.M.I.	·44″	·05″	Paper	100	37·0	$5{\cdot}5 f^{\cdot 68}$	14·0%
HENLEY	D.F.	·5″	·25″	Paper	97	37·0	$1{\cdot}08 f^{\cdot 53}$	9%

from copper tubing. The outer tube is usually supported at frequent intervals a short distance above ground on iron stakes, and special boxes are used to negotiate bends and junctions and to provide for expansion. The general layout of a feeder for a broadside array is shown in Fig. 79, which shows one bay. In early types of feeder, expansion of the outer tube is allowed for by the insertion of corrugated diaphragm - connectors at intervals, but owing to the electrical discontinuity this creates, the expansion of outer as with inner is allowed for now by the use of sliding sleeves. Several sizes of rigid feeder are available, the largest having an outer diameter of 5″, and being capable of handling powers up to 100 kW. at the highest frequencies, the losses in such feeder being shown in Fig. 76.

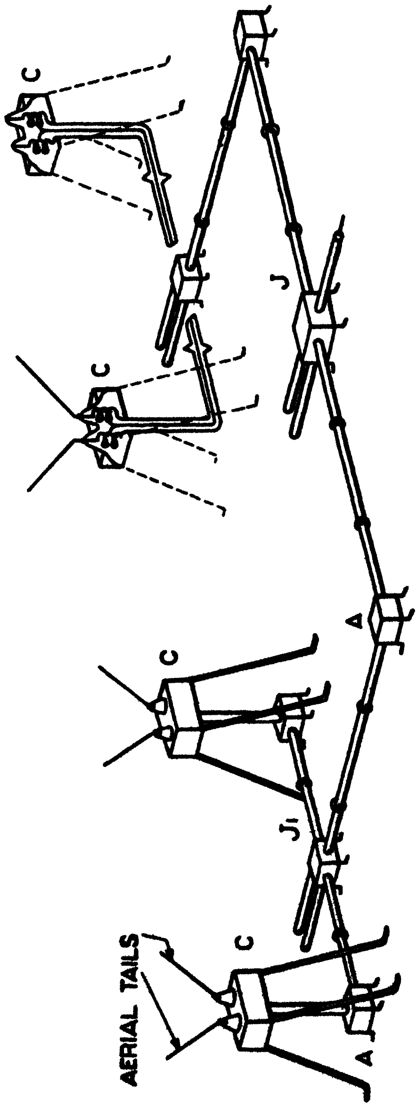


FIGURE 79.

It is important to realise that the maximum potential gradient across the supporting insulator in a concentric tube feeder occurs at the surface of the inner conductor. This means that annular insulators must fit very tightly to the inner conductor or else the thin layer of air between conductor and insulator—since it has a lower dielectric constant than the insulator—will take

the greater part of the voltage between the conductors and on high-power lines a breakdown may take place here.

Concentric tube feeders are sometimes designed to be buried below the ground to reduce expansion troubles and give more perfect screening, and in certain cases arrangements are made to pump in dry air or an inert gas under pressure, in order to exclude moisture and maintain the characteristics of the feeder constant under varying atmospheric conditions. Whether such a refinement is necessary is doubtful, as the experience of Messrs. Cable and Wireless over a large number of years with ordinary concentric tube feeders above ground is that they give no trouble and that their maintenance costs are practically nil, except for the painting of the supporting irons.

Flexible and Semi-Flexible Concentric Feeders. In the last few years a variety of flexible and semi-flexible feeder cables have been developed, the former for receiver work and the latter of sizes large enough for transmission, or for reception where very low loss indeed is required. The advantages of the larger size over rigid types are that they can be handled like ordinary electric power cables and the jointing involves much the same technique, whilst the overall characteristics tend to be more free of discontinuities.

Two types, developed jointly by the Marconi Co. and the Telegraph Construction & Maintenance Company, and manufactured by the latter firm, consist of a solid copper wire as central conductor spaced concentrically within a lead outer tube by means of spacing discs made from a synthetic resin, known as Trolitul, placed about every 3″ apart. Trolitul has a high dielectric constant and a low power factor and has the advantage of not being hygroscopic. The spacers are assembled on the inner copper wire and the lead extruded over the discs from a lead press, and although there is no practical limit to the length that can be built in one operation, it is usual to manufacture such cable in lengths up to 500 metres. After sheathing, the cable is armoured in a normal way with steel tape and jute covering and even the largest size can be rolled on to drums for transport and installation direct into trenches. The installation of the cables follows power cable technique. Thus, for jointing, the centre wire will be scarfed, soldered, bound with

fine wire and resoldered, and the outer lead will be joined with a lead sleeve with wiped joints, and where the joint is exposed the join will be enclosed in a compound-filled cast iron joint box, as shown in Fig. 80. For the termination of such a cable a special end-fixing gland has been designed as shown in Fig. 81.

Although for a given diameter of outer, the lead outer is not quite so efficient as one of copper, ease of manufacture and installation are strong recommendations for its use on low powers.

Smaller and quite flexible cables for receiving work are also made in various types. In these, the insulation is continuous and hence losses are greater and velocity of propagation less. In Table III particulars of a number of cables are given.

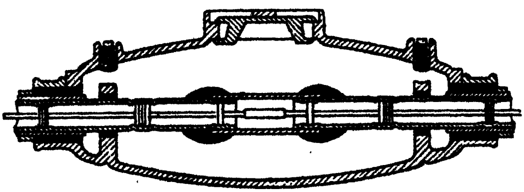


FIGURE 80.

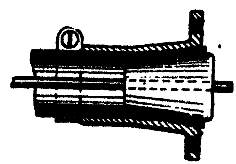


FIGURE 81.

Parallel-Wire Lines. The general construction of an open wire feeder for high-frequency work is similar to that used for telephone lines, except the lines are usually hung from a cross arm on long rod insulators having small dielectric loss, and the longest possible spans will be adopted. The lines are usually hung on poles about 10′ high and the gauge of wire and spacing will be determined, of course, by the power that is to be handled, and where larger powers are used it will be necessary to equip the insulators with corona rings. A gauge of wire of 6 S.W.G., about the largest that can be handled conveniently, is safe for 50 kW. modulated, 100%.

Parallel-wire lines must be carefully balanced both to avoid radiation and prevent losses, not only as regards correct termination but to ensure that each line has the same current. It is frequently found when exploring a line that each line shows evidence of a stationary wave and that the stationary waves are displaced spatially to one another along the line.

This is not necessarily due to an unbalanced load, but may also be caused by unbalanced reflections at supporting insulators.

The efficiency of an open wire feeder is as high as the concentric feeder and Sterba found the measured loss to vary with frequency in much the same way as the calculated, but the measured were much greater, the tests being carried out with a resistance termination equal to the R_0 of the feeder. The B.B.C. made similar tests and found a large variation of input impedance against frequency with a supposedly correctly terminated feeder.

Walmsley gave some results obtained by the British Post Office which clearly showed that large radiation losses will be experienced when feeders are unbalanced, the losses being easily doubled in some cases.

Four-Wire Parallel Wire Feeders. Four-wire feeder lines may be constructed with the go and return pairs disposed at opposite sides of a rectangle, or diagonally. The former arrangement gives a lower surge impedance (see appendix III), but does not make such an easy mechanical lay-out. It is clear that for large powers the four-wire line has the advantage that the conductor size is increased and the characteristic impedance lowered. This means the line voltage for a given power is lower than for the twin wire, the insulator losses are reduced, and because of the mechanical layout of the wires there are no discontinuities and it is found that such feeders are more easy to match up than the twin, and a four-wire feeder with No. 6 gauge wire will handle 100 kW. In Appendix III is shown the calculation for the surge impedance of four-wire lines giving curves for different spacings.

The Measurement of Stationary Waves. In the testing and balancing of parallel wire feeders it is common practice to measure the relative values of the stationary wave maxima and minima, as a knowledge of these enables one to design very quickly one form of correct matching circuit for coupling the feeder to a load circuit (see p. 194).

A number of instruments have been devised for the purpose of measuring stationary waves, the most common being an instrument to run along one wire of the feeder line, and consisting simply of a hot-wire ammeter bridging a portion of the line as indicated in Fig. 82. Results with an instrument of

this kind can be fairly consistent and accurate provided care is taken to clean the wires thoroughly and make sure that good contact is everywhere maintained, that the instrument is kept in the same relative position for all measurements, and that body effects are avoided.

A rather better type of instrument is shown in Fig. 83 which is designed to be free of all contacts, but at the same time picks up energy only through electro-magnetic induction. This is accomplished by designing the pick-up coil in two halves cross-connected as shown in Fig. 83, the output from the coils

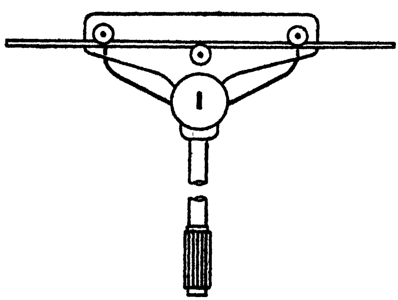


FIGURE 82.

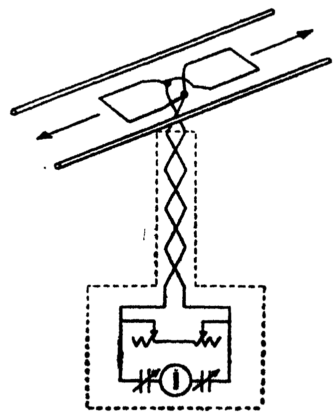


FIGURE 83.

being connected through a screened feeder to the tuned circuit and indicating meter. The whole instrument is carried on an insulating slider (not shown) which locates the coil exactly midway between the lines, and it has been found that extremely accurate results can be obtained.

Summary of Comparison Between Types of Feeders. There is a divergence of opinion regarding the overall merits of concentric-tube and open wire feeders. There is but little difference, if any, in the efficiencies of the two types and the pros and cons are bound up in the economic problem. The rigid concentric-tube type is much more expensive in first costs and to instal and it is not therefore of use for temporary installations. It is non-radiating even if not properly balanced, its characteristics do not change with varying weather conditions,

it is not affected by lightning discharges, and since the outer conductor is earthed the feeder can be installed close to or even buried in the ground, out of the way. The chief disadvantage of this type apart from its cost, is that it does not lend itself to coupling to a balanced type of aerial or array system and in the event of trouble occurring, by a flash over say, it is more difficult to locate and repair.

The chief advantage of the open wire feeder for large powers is its much smaller prime cost and where alterations have to be made, the greater ease of changing feeder runs, and it is much

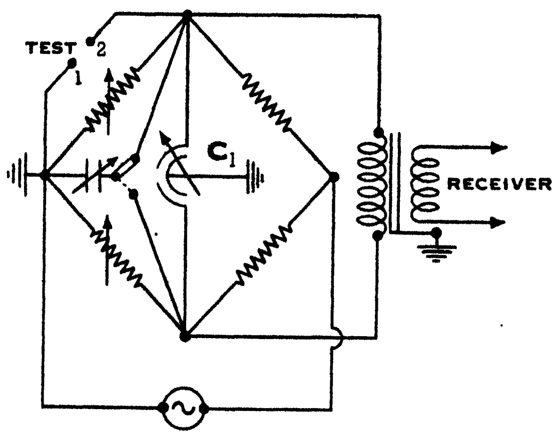


FIGURE 84.

more suited for coupling to the balanced type of aerial, and faults can be traced very quickly. Its disadvantage is its tendency to radiate, it is affected by the weather and it is liable to be struck by lightning discharges.

Impedance Measurements on Feeders. Measurements of feeder and load impedances can be made either by using bridge methods, or by the use of resonant circuits specially adapted for the purpose. The latter have been in use for many years and are most useful for ordinary field work, and the former give accurate and consistent results in expert hands.

Radio-Frequency Impedance Bridge. A balanced bridge (developed by the Marconi Co.) suitable for measurements up to 40 megacycles is shown schematically in Fig. 84, and is designed for use with an oscillator for providing the necessary input

voltage at the desired frequency, and a receiver for balancing purposes, the connections of the oscillator, bridge, and receiver being indicated. It is essential that all three units should each be perfectly screened themselves, and only short screened connecting leads used.

The bridge is designed to cover measurements over a wide range of impedance values, a most essential feature when dealing with any type of aerial system, as will be seen in the next chapter. It is of the admittance type, that is to say, it is balanced by obtaining equality of admittances between the arms of the bridge by the addition of a condenser and resistance, in parallel either with the unknown reactance or with the corresponding balancing arm.

Since the unknown reactance may be either inductive or capacitative, the condenser will need to be placed in parallel with the unknown should this be inductive, but will be put in parallel with the variable arm if the unknown is capacitative.

Having found the values of parallel reactance and resistance which give a balance, these are converted to the usual series equivalents by the formula given on p. 144.

The bridge is first balanced with the test terminals on open circuit, by the differential condenser C_1, the setting of which will vary somewhat for different frequencies. The unknown circuit is then connected to the test terminals, and (having first found whether it is inductive or capacitative) the balance point is found by adjusting condenser and resistances.

Measurement of Feeder Properties by Resonant Circuit. When measurements are being made at very high frequencies it will usually be found simpler to use resonant circuit methods rather than the bridge methods employed at telephone frequencies. This is because, provided that the circuit is well arranged, the inevitable stray capacities do little more than reduce the value of tuning capacity needed, whilst such stray capacities may be extremely troublesome in a bridge network.

A very useful piece of apparatus, therefore, is a resonant circuit consisting of a good coil and condenser together with a valve voltmeter suitable for high frequencies. A stable screened oscillator is required, having sufficient output so that only very weak coupling to the resonant circuit is necessary.

Arrangements should be made so that the circuit may be either (*a*) or (*b*) in Fig. 85.

Such a circuit is frequently called an "impedance matcher."

A large impedance would be measured by means of (*a*) and a low impedance by means of (*b*) and would be connected across terminals 1–2. The reactance of the apparatus under test is found from the change of C necessary in order to re-tune when it is inserted. The resistance is found by adjusting R until the voltmeter has the same reading for either position of S (tuning to resonance each reading).

In showing how to deduce feeder attenuation, etc., from measurements, it will be necessary to use the full telephone line equations. Readers not familiar with these are referred to Appendix IV where they are set out.

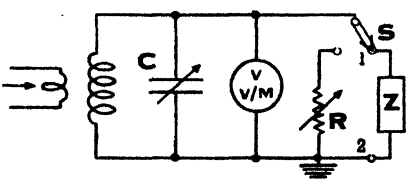


FIGURE 85a.

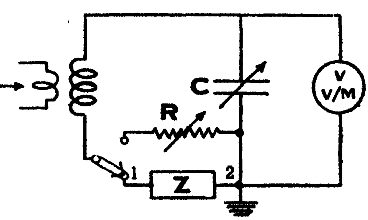


FIGURE 85b.

The characteristic impedance and attenuation constant of a feeder can be determined by using a quarter-wavelength of the feeder, in the following way. Connect the feeder as Z in Fig. 85 (*b*) (the far-end being open-circuited) and adjust oscillator frequency slightly until no change of C is required when S is changed. The feeder is now exactly $\frac{\lambda}{4}$ long and the impedance placed across the tuned circuit is therefore $Z_{oe} = Z_o \tanh al$, the value of which can be determined by adjusting R. The far end is now short-circuited and Z_{se} determined, but the circuit (*a*) will be necessary as $Z_{se} = Z_o \coth al$ and will be large. Then $Z_o = \sqrt{Z_{oe}\,Z_{se}}$ and $\tanh^2 al = \frac{Z_{oe}}{Z_{se}}$. The measurements may also be made at harmonic frequencies where similar conditions will hold.

Another method for finding a consists of measuring the voltages at each end of a quarter-wavelength of feeder, open-circuited.

Then $E_x = E_g \cdot \dfrac{\cosh P(l-x)}{\cosh Pl}$ and this becomes $\dfrac{E_g}{E_l} = j \sinh al$ if feeder is an odd quarter-wave long. The voltage at the far end will, in general, be several times greater than that at the input end and if $\dfrac{E_g}{E_l}$ is less than 0·1 we can consider $al = \dfrac{E_g}{E_l}$.

It has been found possible to estimate the conductor and dielectric losses separately by measurements on a short length of feeder. If a short length of open-circuited feeder is placed across a very good tuned circuit, then we can assume that the conductor loss due to the small current which will enter the feeder is negligible and that the resistance thrown across the tuned circuit (which can be determined by a resonance curve) is entirely due to dielectric loss.

When measuring a series impedance such as $r + jx$, it may happen that the most convenient apparatus available will measure parallel components R_p and X_p say. If we insert a quarter-wave feeder between the impedance to be measured and the measuring circuit, then $r = \dfrac{R_0^2}{R_p}$ and $x = -\dfrac{R_0^2}{X_p}$, as shown on page 145.

Suppose a feeder to be perfectly uniform, that is, its electrical properties are quite uniformly distributed along its length. If it is now correctly terminated and the "sending-end" impedance measured for a range of frequencies, this will be found to be constant and equal to the terminating resistance. If, however, there are irregularities in the electrical properties due to insulators, bends, joints, etc., then the sending-end impedance will have maximum and minimum values as the frequency is varied. It is possible, in some cases, to find the position of the irregularity from these tests.

In adjusting a complete feeder system such as that for a Marconi broadside array, the "impedance matcher" would be used in the following way.

We would commence with one aerial pair, inserting the impedance meter in the feeder at the first transformer box (J_1), (Fig. 79), (transformer and the other branch feeder removed). By adjusting aerial tappings and feeder tappings on the aerial coils the correct termination can be obtained. All aerial coils are then set to the same adjustments, as the aerials, etc., are identical, and, in fact, the adjustments must be exactly the same to ensure that all aerials are supplied with current in the same phase. The impedance matcher is now removed and placed at J and the transformer coil at J_1, fitted. By adjustment of the tap on the transformer coil, the branched feeders are made to give an equivalent termination of R_o. The transformers at J and all others can then all be set to the same adjustment, and the impedance matcher is then moved back to the next junction box, and so step by step to the transmitter.

All these adjustments are made with a low power output from the transmitter, but when the impedance matcher (now mounted at the transmitter end of the feeder) indicates that the whole system is correctly terminated, the matcher may be removed and full power put on the transmitter. By adjusting the coupling between transmitter and feeder the full output from the transmitter is obtained. To obtain a final check under full power conditions three ammeters can be connected in the feeder, the spacing between the two outer meters being somewhat less than half a wavelength. In some installations a length of feeder is bent into a U shape in order that all three meters may for convenience be inside the transmitter room. It is clear that if all the meters read the same value of current, no stationary wave is left in the line, and the power going to the main feeder is $I^2 R_o$. Care must be taken that the meters are not left in circuit when preliminary adjustments are being made, as any stationary wave set up might burn them out.

Special Precautions in Feeders for Television. In most applications of short and ultra-short waves, the modulation frequency is a very small percentage of the carrier frequency and therefore any ordinary resonant circuits, coupling arrangements, feeders, etc., will work equally well over the frequency bandwidth involved in the transmission.

Television forms a complete exception to this rule and special care is necessary in the design of the feeders and aerial system.

The aerial system will show a considerable change in impedance over the frequency band involved. If an approximately $\frac{\lambda}{4}$ line is used as transformer to couple aerial to feeder, the change in its behaviour due to its different electrical length at the different frequencies can be made to compensate for the aerial changes and keep the main feeder almost correctly terminated for all the sideband frequencies.

Because of the shorter wavelength, insulator spacing will become important. All irregularities such as angle boxes will have to be compensated for by very small condensers at suitable points.

If the feeder is long, any reflected wave due to mis-matching or feeder irregularities will provide a spurious signal a sufficient time after the true signal to give multiple images in the received picture.

Dielectric Wave-Guides. Lord Rayleigh showed, in 1897, that it should be possible to transmit electromagnetic waves through the dielectric contained in a hollow conducting cylinder, without the presence of a central conductor.

Such waves are only possible if the diameter of the cylinder is of the same order as the wavelength used (with an air dielectric) and hence very high frequencies are necessary with cylinders of practicable dimensions. In consequence, experimental work upon the subject has only been undertaken seriously in recent years, when the necessary sources of very high frequencies have become available.

Southworth first experimented with water in a cylinder, because the greater the dielectric constant, the lower is the frequency which can be employed. At high frequencies, however, the dielectric loss (and hence the attenuation) is very great, if any dielectric other than air is used. In recent work wavelengths as short as 0·25 cm. have been used.

Four different types of wave have been produced, usually termed E_0, E_1, H_0 and H_1 waves, each characterised by a different distribution of the electric and magnetic fields. In all cases there is a longitudal component of either E or H, as

compared with the purely transverse waves propagated along a concentric tube cable.

Consider a source of one of these types of waves placed at one end of the wave-guide and a suitable detector placed at the other. If the frequency of the source is gradually raised from a low value, the response of the detector varies with frequency in the same way as a high-pass filter response-curve. That is, there is at first no response whatever, then at a definite frequency corresponding to the cut-off frequency of a filter (which is lowest for the H_1 type of wave) an output is obtained and as the frequency is raised further the response continues to increase because the attenuation becomes less.

At the very high frequencies for which it is practicable to use dielectric guides, they produce less attenuation than concentric lines would do and are, of course, simpler in their construction. It is probable, therefore, that dielectric guides will be considerably employed as the very high frequencies come more into use.

SELECTED REFERENCES.

The Propagation of Electric Currents. Fleming. Constable.

Telegraphy and Telephony. Mallett. Chapman & Hall.

Short Wave Aerial Systems. Green. W. E.; June, 1928.

Beam Arrays and Transmission Lines. Walmsley. *J.I.E.E.*, Vol. 69; February, 1931.

Transmission Lines for Short Wave Radio Systems. Sterba and Feldman. *P.I.R.E.*, Vol. 20; July, 1932.

Long Feeders for Transmitting Wide Side-bands. Cork and Pawsey. *J.I.E.E.*, Vol. 84; 1939.

Empire Broadcasting Station. Hayes and MacLarty. *J.I.E.E.*, Vol. 85; 1939.

Hyper-frequency Wave Guides. Southworth. *B.S.T.J.*; 1936; p. 284.

Transmission of E.M. Waves in Hollow Tubes. Barrow. *P.I.R.E.*, Vol. 24; October, 1936.

Theoretical Relationships of Dielectric Guides. Clavier. *Elec. Com*; January, 1939.

Experimental Researches on Dielectric Guides. Clavier and Altovsky. *Elec. Com.*; July, 1939.

CHAPTER VII

AERIALS

In the same way that wireless waves fall naturally into two broad divisions, long and short, so aerials fall into two classes, according to whether they are to be used for long or short waves. With long waves the length of wire is but a fraction of the wavelength, and the aerial must therefore be "loaded" with inductance until it is equivalent to a quarter wavelength aerial. The usual long wave aerial consists of a vertical portion which supplies the useful radiation and a flat portion, or "roof," the capacity of which serves to increase the current in the vertical and make its distribution uniform, thereby increasing the radiation.

In the case of short waves, however, it is practicable to construct an aerial which may be many wavelengths long, and it will usually consist of a straight wire. The "roof" may be dispensed with, because if the aerial is a quarter wave or more long there will be no difficulty in producing a large stationary wave of current without having an undue voltage at the high potential end. One of the advantages of short waves is that a small and simple aerial will accept a large output from a transmitter and radiate a great proportion of it. The short wave aerial is, therefore, much more efficient as a generator of electromagnetic waves, even though a good proportion of the radiation (how much has not yet been determined with any certainty) is wasted in the neighbourhood of the aerial, due to earth losses, etc.

The tuning of a long wave aerial is a relatively simple matter, because the aerial length will always be much less than $\frac{\lambda}{4}$, and hence we have only to "load" it with inductance to tune to any wavelength within the long wave band. Thus a large "tuning range" can be covered, although, as the dimensions

of the aerial become a smaller and smaller proportion of the wavelength being used, the radiated power for a given input power will decrease.

As we shall show later, the tuning of a short wave aerial, however, is a more complicated matter, because its behaviour is subject to sudden disconcerting changes as the applied frequency is changed, and hence considerable difficulty may be experienced in "fitting" the aerial to a circuit unless proper precautions are taken.

In general, we have two distinct types of short wave aerials—those which are designed to give the greatest possible efficiency at one given wavelength and are not usually suitable for covering any wave range, and those which cover a wide wave range and sacrifice efficiency to achieve this.

Distribution of Current, etc. Before discussing the radiating properties of the various types of aerials, we will consider the behaviour of an aerial as a circuit. In particular, we require to know the current distribution along it, and its impedance measured between the points at which it is to be fed (usually between the base and earth). A knowledge of this impedance is necessary in order that we may arrange suitable coupling for transferring energy to or from the aerial when it is connected to transmitter or receiver.

Although electromagnetic waves are produced whenever a varying current flows along a conductor, and certain short wave aerial systems do operate by virtue of a travelling wave along a correctly terminated wire, radiation is usually associated with a system in which stationary waves are present, and we will consider the operation of aerials working under such conditions. In order to radiate at all efficiently it is necessary that the dimensions of the circuit should be comparable with the wavelength, and the simplest method of getting a large current into such a circuit is to produce stationary waves.

Radiation is greatest from those parts of a circuit where the varying current is greatest (unless influenced by adjacent conductors) and if the length of the system is long compared with the wavelength, several modes of oscillation are possible.

Whatever the length or form of aerial, it is clear there can be no current at the end remote from the earth, but it may be possible, by suitably arranging a circuit between the lower end

of the aerial and earth, to produce stationary waves such that either a voltage or a current node is produced at the lower end.

Consider a vertical wire of length l, fed from a generator of frequency f, as shown in Fig. 86. If we induce current in it such that the wire is approximately one-quarter of the wave-length in free space corresponding to f, then by a small variation of the length of wire (or of the applied frequency) we shall obtain

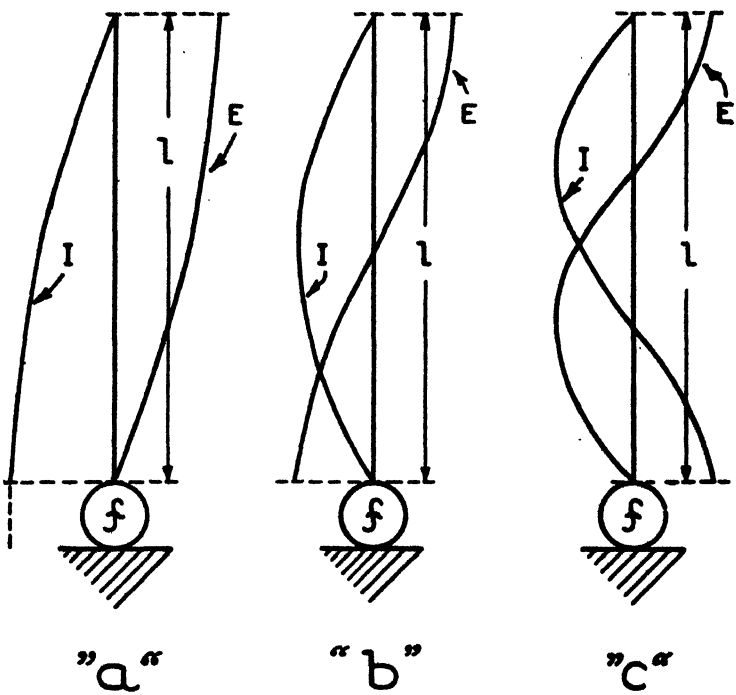


FIGURE 86.

a resonance at which the current is a maximum at the base of the wire. This resonance point is called the natural wavelength, and the distribution of current and voltage will be as shown in Fig. 86a, from which we can observe that the wire is carrying a $\lambda/4$ stationary wave. The actual length of the wire is less than a quarter of the wavelength, and more nearly $\lambda/4{\cdot}5$, because the capacity and inductance are not uniform, and for this reason the current and voltage distribution is not truly sinusoidal.

If, now, we increase the length of wire to $\lambda/2$ approximately (or, with the same wire, double the applied frequency), another resonance point will be found (the natural wavelength of an unearthed wire), but this time a voltage antinode occurs at the base end of the wire, and the distribution of E and I is as shown in Fig. 86b. As before, the exact length of wire to give resonance is not equal to $\cdot 5\lambda$, but more nearly $\cdot 47\lambda$, the exact figure being largely dependent on the proximity of other objects and the method of supporting the free ends.

A further increase of length to three-quarters of a wavelength (or increase of frequency to $3f$) shows another resonance with maximum current at the base and a distribution of E and I as shown in Fig. 86c. In fact a whole series of resonance points will be obtained, the tuning at the odd quarter wavelengths producing maximum *current* at the base, and tuning at the half wavelengths maximum *voltage* at the base. Such an aerial operating at a multiple frequency is called a harmonic aerial, the exact length of aerial in terms of the harmonic frequency becoming more nearly the exact theoretical figure for the higher harmonics.

If the aerial had uniformly distributed inductance and capacity and no resistance or radiation, then its behaviour would be identical with that of the open-ended feeder discussed in the previous chapter, and hence we can get useful approximate ideas by applying the feeder analysis to our aerial problem.

It was shown that the reactance of an open-ended feeder was given by

$$X = -j\sqrt{\frac{L}{C}}\cot \omega h\sqrt{LC}$$

where h is the length of feeder, or, in this case, the height of the aerial. Maximum current at the foot of the aerial will evidently occur when this is zero, that is when

$$\omega h\sqrt{LC} = \frac{\pi}{2} \text{ or } \left(n + \frac{1}{2}\right)\pi,$$

n being an integer.

In the case of a straight wire in clear surroundings, $\dfrac{1}{\sqrt{LC}} = c$, the velocity of light;

$$c = \frac{\omega}{2\pi}\cdot\lambda$$

Hence $\omega h \dfrac{2\pi}{\omega\lambda} = \dfrac{\pi}{2}$ or $h = \dfrac{\lambda}{4}$

as we have already seen to be the case with the assumptions made. Professor Howe ("Wireless Year Book," 1917) finds that

$$L = 2\left(\log_e \frac{2h}{d} - 1\right) 10^{-9} \text{ henries,}$$

is a close approximation for the inductance per cm. of a vertical wire near earth and

$$C = \frac{1}{\left(2 \log_e \dfrac{2h}{d} - 1\right)} \times \frac{1}{9 \times 10^{11}} \text{ farads}$$

for the capacity per cm., where d is the diameter of wire in cms., and h the length, also in cms.

It is of interest to observe that the characteristic resistance of such an aerial will be

$$R_o = \sqrt{\frac{L}{C}} = 60\left(\log_e \frac{2h}{d} - 1\right).$$

For instance, an aerial 15 metres long, 0·25 cm. diameter, will have a characteristic resistance of

$$R_o = 60\left(\log_e \frac{3{,}000}{\cdot 25} - 1\right) = 510 \text{ ohms.}$$

If we have a wire of definite length and work out its constants as above, then its frequency reactance curve will be as shown in Fig. 87, this curve showing the multiple tuning points just discussed. Thus at every odd quarter wavelength its reactance measured between the base end and earth (in future called base reactance) is zero, and (since the aerial has been assumed resistanceless) the base resistance is zero. For every even harmonic the base resistance is infinity and hence any circuit used to feed the aerial between the base and earth must be capable of impedance variation between wide limits, if it is to match the aerial.

For instance, if an aerial is at exactly quarter wave resonance (or odd multiples of this) with the applied frequency, then its base reactance is zero, and a suitable coupling would be a series tuned circuit of L_1, C_1, which will also have zero reactance, as shown in Fig. 88a. If the aerial is somewhere near quarter

wave resonance, but not exactly in tune, then the base circuit should also be detuned until the sum of the reactances is zero, that is, with the assumptions made :

$$\sqrt{\frac{L}{C}}\cot \omega h\sqrt{LC} = \omega L_1 - \frac{1}{\omega C_1}$$

This equation can be solved by plotting the reactance curves of the aerial and of the tuning circuit, but the result would be

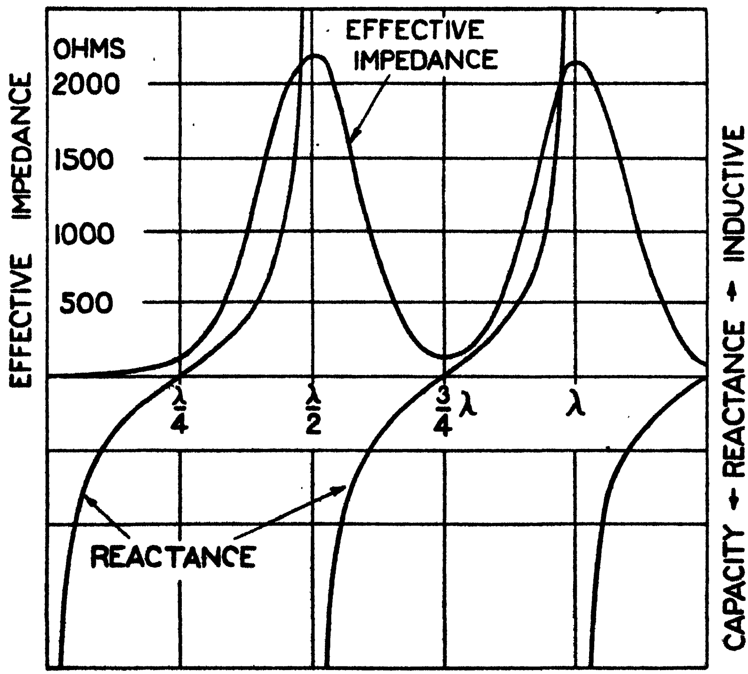


FIGURE 78.

very approximate, due to the assumptions made regarding the aerial reactance, and in practice the correct adjustment of an aerial is easily found by varying the tuning circuit until maximum current is obtained in the earth connection.

At the $\frac{1}{2}\lambda$, λ, $\frac{3}{2}\lambda$, tuning points, the aerial acts as a parallel resonant circuit, its base reactance is zero, but base resistance infinity and the only circuit to which we can couple it efficiently is, therefore, one of parallel inductance and capacity, whose resistance at resonance is also infinite, the arrangement being shown in Fig. 88b.

In the same way as before, if the aerial is nearly, but not quite a half wave (or multiple thereof), the parallel circuit needs to be de-tuned. In this case no convenient aerial current reading is possible, since at the base end the current is only a small feed current, and correct adjustment is best found for transmitting work by observing the load which the aerial puts upon the transmitter.

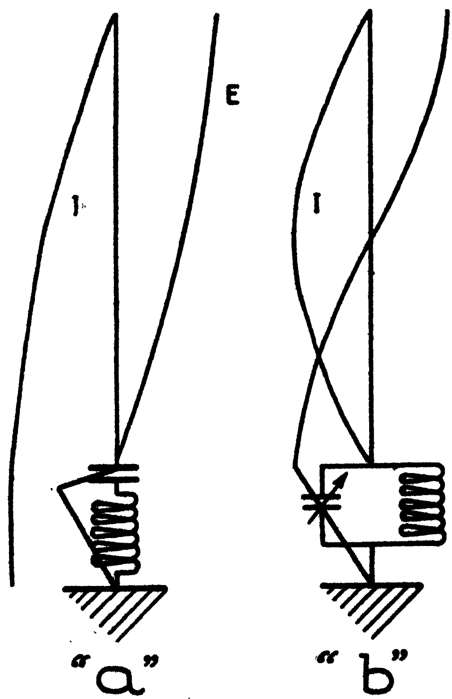


FIGURE 88.

We will now consider how resistance modifies the above cases. The aerial is not a loss-free feeder, but if a quarter wavelength or more, has a considerable radiation resistance as well as losses, which vary with its mode of oscillation, and its proximity to earth.

The aerial disposes of energy in the following ways:

(1) Radiation of electromagnetic waves.

(2) Losses in the earth and due to eddy currents in neighbouring conductors.

(3) Losses in aerial wire and in insulators, etc.

Since these losses are proportional to the square of the current, they can be replaced in our consideration of the aerial as a circuit, by an equivalent entirely fictitious resistance located at any convenient point in the aerial. Since, however, the aerial is carrying a stationary wave, the value obtained for this resistance will vary, being least at current antinodes and increasing to high values near the current nodes.

To avoid ambiguity, therefore, it is necessary to specify the point at which this effective resistance is placed, the current antinode always being chosen.

Actual values for effective resistance are difficult to calculate (or measure) at short wavelengths, but for a quarter wave aerial a value of 50 ohms is normal (referred to the current antinode at earth, and for a half-wave aerial 150 ohms (referred to the current antinode—at centre in this case).

In the quarter-wave case, since the effective resistance is taken at the base of the aerial, the base resistance is also about 50 ohms.

In the half wave case the feed is at the base of the aerial, where the voltage is a maximum, whilst the effective resistance R is referred to the centre of the aerial where the current is a maximum. Hence the base resistance will not be R but some higher value.

We can get at the base resistance of a half-wave aerial approximately from the following considerations. Let us try to construct a closed resonant circuit composed of *concentrated* inductance, capacity and resistance which shall behave in the same way as the half wave aerial having *distributed* inductance and capacity. Note that the aerial and the closed circuit will only behave in the same way for the exact frequency which produces half wave resonance of the aerial. Note also that the concentrated inductance and capacity we are going to use in our closed circuit represent values which could not actually be measured in the aerial, the only measurable quantity being the base resistance which we are trying to calculate.

Suppose that the closed circuit carries in all its parts a current I equal to the maximum current in the aerial—that is, the current at the centre. If the aerial (of height h cms.) has a uniformly distributed inductance of L per cm., then its total effective inductance would be Lh, if it carried the

current I throughout. To allow for the actual distribution we should use the mean value of the current which is $I.\frac{2}{\pi}$ (for a half-sine wave) and hence the total effective inductance L_e is $Lh.\frac{2}{\pi}$ and this is the value of the inductance in our equivalent closed circuit, as shown in Fig. 89.

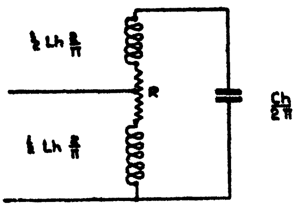


FIGURE 89.

Dealing with the capacity of the aerial in a similar way, we shall need in this case to consider the two quarter-waves separately, because the voltages are of opposite sign. If C is the capacity per unit length, $\frac{Ch}{2}$ would be the effective capacity if the voltage was uniform but $\frac{Ch}{2}\cdot\frac{2}{\pi}$ if the voltage is distributed in a quarter-sine wave. The two quarter waves are in series and therefore the total effective capacity C_e is $\frac{Ch}{2\pi}$.

The half-wave aerial is at earth potential at its centre point and, therefore, the base impedance is really measured between one end and the centre of the aerial, and the arrangement of the equivalent circuit is as shown. It is seen to be a parallel resonant circuit, tapped half way down and its equivalent resistance at resonance is therefore

$$R_e = \frac{\omega^2}{R}\left(\frac{L_e}{2}\right)^2 = \frac{\omega^2 L_e^2}{4R}$$

Since $\omega^2 L_e C_e = 1$, this may be written $\frac{L_e}{4RC}$, or in terms of L and C,

$$R_e = \frac{Lh\frac{2}{\pi}}{4R\frac{Ch}{2\pi}} = \frac{L}{RC}$$

If we take as an example a half wave aerial 1500 cms. high and of 0·25 cm. diameter, then, from the formulæ on p. 170,

$$\frac{L}{C} = 4\left(\log_e \frac{3000}{\cdot 25} - 1\right)^2 \times 9 \times 10^2 = 258{,}000$$

and if R is 160 ohms then the base resistance, $\frac{L}{CR} = 1{,}600$ ohms.

It will be seen that an increase in the diameter of wire used will decrease L and increase C. It is reasonable to assume R to be unaffected by the change because R is mainly radiation resistance (on short waves) and the conductor resistance will have negligible effect.

In consequence, the base resistance decreases if the wire diameter is increased, but the effect is not very marked unless a cage aerial or tubular conductor is employed. For instance, doubling the diameter in the example given would only reduce R_e to 1,300 ohms, but for a tube of 2 cms. diameter R_e would become 895 ohms.

Further, the Q value of the equivalent circuit is reduced if L is reduced, and this means that the frequency/response curve is broader, or the aerial considered as a tuned circuit is less selective.

This feature is useful in television aerial design. Thus the Alexandra Palace aerials consist of triangular cages of 15 inch sides, and for television reception it is common practice to use a tubular half-wave aerial which is good both from an electrical and mechanical point of view.

Thus, in the case of an aerial having resistive loading due to radiation and losses, the value of impedance varies over extremely wide limits as the frequency is changed (see Fig. 87), and it can be seen that to tune an aerial to wavelengths between the natural resonance points necessitates not merely the alteration of reactance of tuning circuit, but an essential alteration of the type of coupling circuit.

In the case where one wavelength only is required, we do not need to have a variable tuning circuit. An alternative method of feeding the aerial can be arranged, therefore, namely through a length of wire (called a tail) and a coil, as shown in Fig. 90a. For if the curve of Fig. 87 be referred

to, it is clear than an aerial longer than $\frac{\lambda}{2}$, but less than $\frac{3}{4}\lambda$ has a capacitive reactance between its base and earth, and thus between these limits, an inductive loading will be required to tune to resonance.

The conditions very near the half wave point are so influenced by high base resistance, that it is found very difficult to tune an aerial only just a fraction longer than $\frac{\lambda}{2}$ by inductive loading, and it is customary to add a tail between the aerial proper and the loading coil, as shown in Fig. 90a; curve 90b shows the stationary wave on aerial tail and coil, the current in the latter being uniform. This tail should be carried horizontal to earth and preferably doubled back on itself so as to reduce radiation therefrom, and its length should not be less than $\frac{\lambda}{10}$ or more than $\frac{\lambda}{8}$. If the tail is too short it is difficult to tune as previously explained, whereas if it is too long, the base current is higher, more radiation from the tail is experienced, and earth losses are greater.

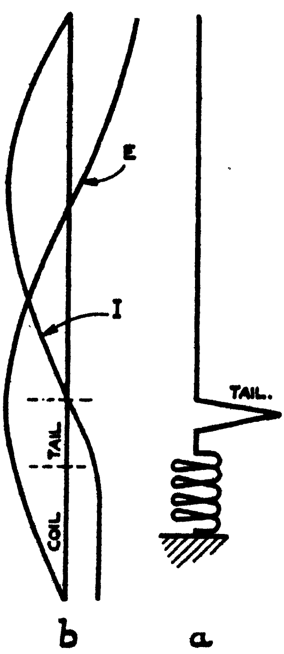


FIGURE 90.

Radiation from an Aerial. The radiation from the simplest radiator—the dipole—has already been discussed in Chapter IV.

Actual aerials are not dipoles because their dimensions are comparable with the wavelength radiated and the current in the various parts is not uniform. In the long-wave case, where the aerial is usually only a fraction of the wavelength, the dipole equations may be used in obtaining field strengths at a distance by inserting an equivalent height to allow for the current distribution. Such a method is not applicable to the short-wave aerial, however, since it may be more than a

wavelength high, will have a much less uniform current distribution and will not have a polar diagram of cosine form. It will be necessary to treat the actual aerial as built up of a number of elementary lengths each having a cosine polar diagram and

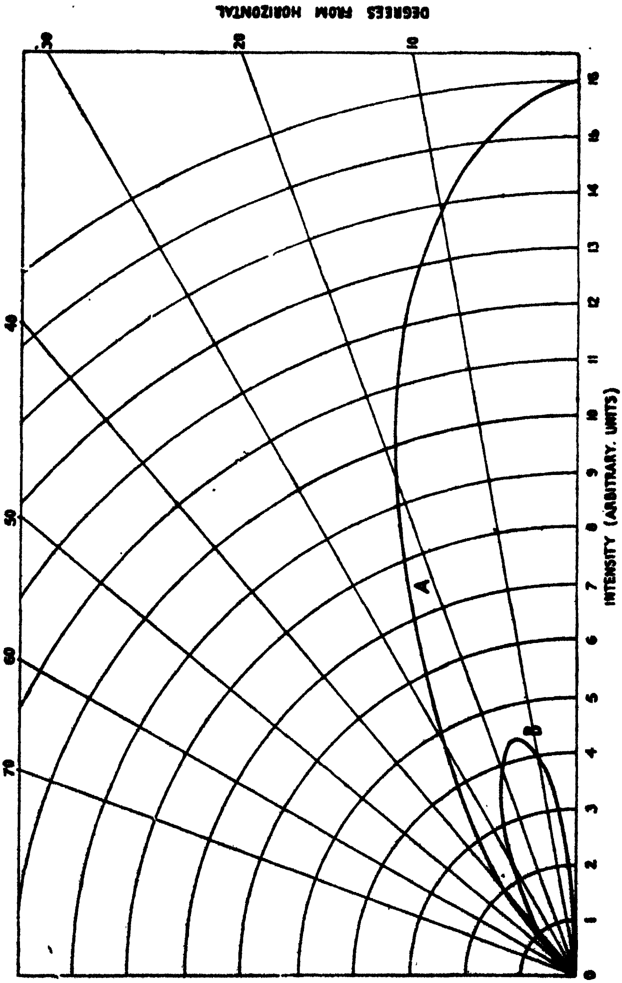


FIGURE 91.

carrying currents appropriate to their position in the aerial.

The effect of the earth will be allowed for by an image aerial, also split up into elements presumed to be carrying currents of magnitude and phase dependent upon earth constants, in the way discussed in Chapter IV.

For example, the zenithal polar diagram for a half wave aerial just above a perfectly conducting earth is shown in Curve *A*, Fig. 91, and is seen to be much sharper than that of a dipole.

From this polar diagram a value for the radiation resistance has been deduced by Ballantine using the method explained on page 177, and a value of 104 ohms obtained. By the same method the radiation resistance of a quarter-wave aerial with its lower end earthed, can be worked out and is found to be 36·6 ohms. Diagrams for a half wave aerial at various heights above a perfectly conducting earth are shown by the dotted curves in Fig. 92.

These results assume a sinusoidal distribution of current which is everywhere in the same time phase—that is, a perfect stationary wave, whereas, since energy is being radiated, the actual current must be a combination of a stationary wave and a travelling, energy-conveying wave.

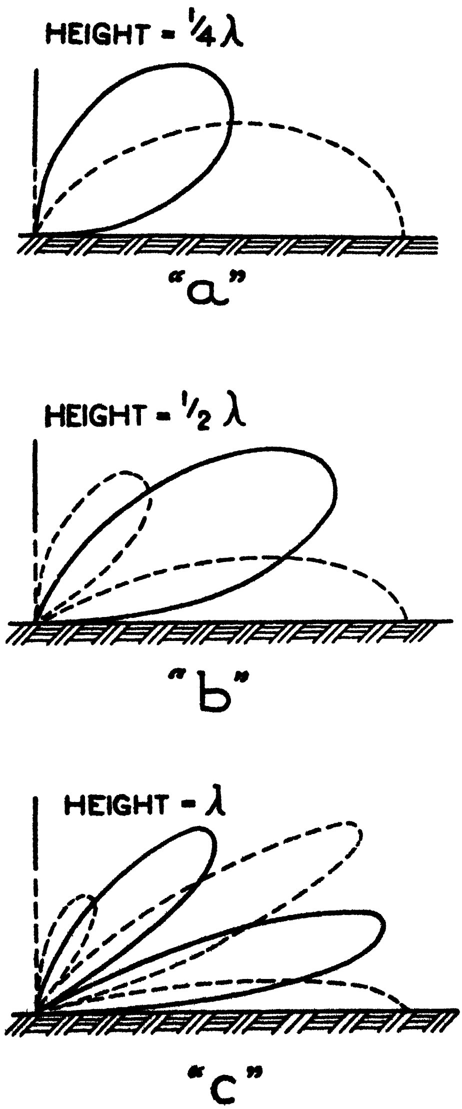


FIGURE 92.

Effect of Imperfectly Conducting Earth on Zenithal Polar Diagram of Aerial. If the finite conductivity and

dielectric constant of the earth is taken into account, the polar diagram for a given aerial is dependent upon frequency. The polar diagram, since it is dependent on the "image" theory which assumes optical reflection at the earth's surface, does not take account of the "surface wave." In other words, the diagram is really only approximate for great distances and ignores energy transmitted along the earth's surface, this being quickly absorbed in the case of short waves.

The polar diagram for a half-wave vertical aerial radiating a 22 metre wave and situated just above the surface of earth having likely values of conductivity and dielectric constant, has been calculated and is shown by curve *B* in Fig. 91, where it is compared with the polar diagram when the earth is perfectly conducting. It will be seen that whereas the diagram, when the earth is considered to be perfectly conducting, has its maximum value in a horizontal direction, when likely earth constants are assumed there is no horizontal radiation at the wavelength considered.

The radiation resistance of an aerial producing a polar diagram such as curve *B*, can be found by comparing the area of curve *B* with the area of curve *A* since the radiation resistance in the latter case is known to be 104 ohms, and the polar diagrams are the same in all horizontal directions. Note that a direct comparison of polar diagram areas does not give a correct result because the field strength at a low angle on the surface of the hemisphere is the same over a zone having considerable area whilst the field strength at a high angle only exists over a zone of small radius. When this difference of area has been allowed for the method shows the radiation resistance to be 31 ohms whilst the total effective resistance of such an aerial has been measured experimentally and found to be about 165 ohms so that its efficiency for long distance communication is about 18%.

Much of the energy which appears to have been lost is probably radiated in the surface wave and theory and experiment agree that raising the aerial greatly improves the efficiency.

It has already been explained that approximate polar diagrams of vertical short-wave aerials are best obtained by considering the image aerial to carry an equal current in anti-

phase for the lower angles of elevation and an equal current in-phase for the higher angles.

Curves for half wave aerials are shown in Fig. 92, in which the full curves are for an anti-phase image and dotted curves for an in-phase image.

It should be clearly understood that on short waves a statement of the radiation resistance of an aerial is not sufficient to enable the merits of the aerial for a given purpose to be assessed. Much more knowledge of its usefulness for a given purpose can be derived from a study of its polar diagram.

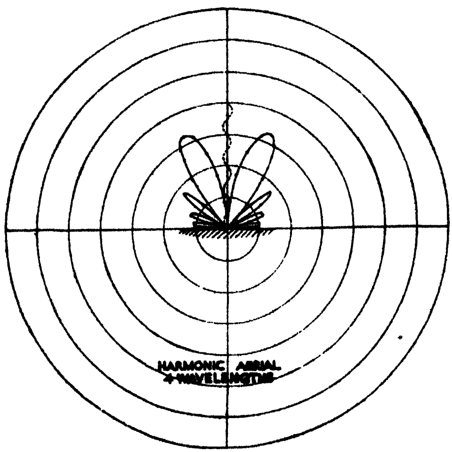


FIGURE 93.

Zenithal Diagram of Harmonic Aerials. We can get increased height merely by making the wire longer and having a number of half-waves on it forming a harmonic aerial, but if we do this we get increased high angle radiation, whilst the radiation from the various half-waves will cancel out at low angles, because adjacent half-waves are in phase opposition The zenithal polar diagram of a four wavelength, harmonic aerial is shown in Fig. 93. But it has already been pointed out (in Chapter V) that it is of considerable advantage when transmitting to concentrate the energy radiated into a sharp " beam " at a low angle because we have only a definite amount of energy available and must use it in the most effective way. When receiving it is equally important to use an aerial to receive over only a small sector having a small angle to earth,

to avoid multiple effects. Thus to get increased height effectively and yet obtain concentration of field at low zenithal angles necessitates a special type of aerial.

Tiered Aerials. To gain this object C. S. Franklin developed the tiered aerial, earlier types suppressing the alternate half-wave radiation, and later types using the alternate

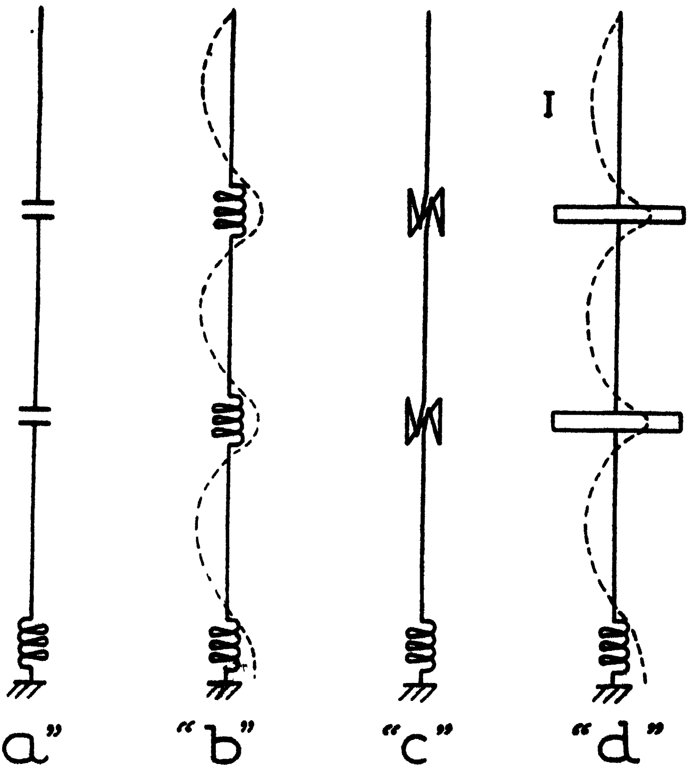


FIGURE 94.

half-wave portions to assist in producing radiation in the required direction.

Franklin's first tiered aerial (1922) comprised a series of half-wave aerials with coupling condensers between each, as shown in Fig. 94a, the small capacity between the ends of the wires being sufficient to transfer energy at the very short waves used. Then followed the phasing-coil type in various forms whose shape was designed "to concentrate alternate half

wavelength portions of the wire within a small space so that there is practically no radiation from these portions," to quote from the Patent Specification. Some of these types are shown in Figs. 94b, 94c, 94d, and on two of them is shown the form of stationary waves produced. In practice the proportions of wire lengths are not half-waves, but empirical rules have been found to give the best results. The length of wire below the lowest phasing unit is extended beyond a half wavelength by a tail of such value that it makes the whole aerial capacitative in reactance in order to make it suitable for fitting to a quarter-wave tuning coil.

Although such types of aerial marked a great advance in the efficiency of short wave aerials, they were not ideal as the abrupt change of characteristic from open aerial to "lumped" inductance-capacity caused reflections at the various points, and resulted in a reduction of radiation from the top portion of the system, this being undesirable. To overcome these defects, the present uniform aerial was produced.

Franklin Uniform Aerial. In order to get the greatest possible concentration of radiation at a low zenithal angle with a given total height of aerial, the ideal arrangement is a vertical wire carrying a uniform current in the same phase. An aerial approximating to the ideal can be made in a variety of ways, one type being indicated in Fig. 95, which shows that each successive half-wave wire is folded back on itself in such a manner that the radiation from its central part assists radiation from adjacent wires. The radiation from the tips of the phasing wire cancels the radiation from the tips of the adjacent wires, but since it is the sections carrying the maximum current which are chiefly productive of radiation, the elimination of radiation from the ends does not matter. By such means we attain almost to the ideal of a uniform current aerial as indicated in the figure, and thus utilise the available height in the most economic manner possible.

The increase in the number of wavelengths of wire for a given linear extension of aerial by such folding results in the power radiated (and of course the power drawn from the transmitter) being greater for the same current than in the case of a half-wave aerial.

If we have high masts and only a given amount of power

available, it may be more advantageous to concentrate this in a few aerials well away from earth rather than in more aerials, some of which are close to earth. This can be done by using a length of single wire feeder of a nearly non-radiating type between the feeder proper and aerial. A type used by Franklin is shown in Fig. 95, and consists of a wire bent back on itself in such a way that in the aggregate the radiation is mainly cancelled.

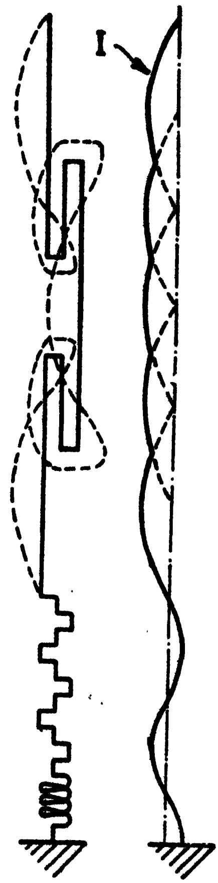


FIGURE 95.

Vertical Polar Diagrams of Tiered Aerials. We can determine approximately the vertical polar diagrams of a system of tiered aerials by the following method, which assumes the aerial to carry a uniform current of the same phase throughout its length. The effect of the earth is taken into account by an image aerial of opposite phase, since we are mainly interested in the low angle radiation.

Consider a vertical aerial of length $l=n\lambda$ (Fig. 96a) with a corresponding image aerial as shown. If we imagine the aerial to be divided into elementary lengths of aerials all carrying a uniform current, then the field produced at a point P, distant from O, at any angle $\theta°$ from the ground, will be determined by the vector sum of the individual fields. If the field due to one element be E, and there are x elements, we have to sum up x vectors, each of which is slightly out of phase with the preceding one, because the distance between each element and P is not the same. These vectors form approximately the arc of a circle $A'B'$, as shown in Fig. 96b, the chord being the resultant, the shape of this arc depending upon the angle being considered and the aerial length.

Now since the arc is the arithmetic sum of the fields its length is constant (say unity) and hence the ratio $\frac{\text{chord}}{\text{arc}}$ determines the relative field at any given direction. Consider

Fig. 96b. The phase angle between the first and last vector equals ϕ radians, the angle at the centre of the circle subtended by the arc. The ratio

$$\frac{\text{chord}}{\text{arc}} = \frac{\sin\frac{\phi}{2}}{\frac{\phi}{2}}$$

Hence, if we find the phase difference between the first and last elements of the aerial in terms of the zenithal angle θ we can obtain the relative field strengths from the above expression.

The extra distance to be covered by the wave from the lowest element is OC (Fig. 96a). If this distance is λ, the phase difference would be 2π radians, hence the actual phase difference ϕ is given by

$$\frac{2\pi}{\lambda} \cdot OC$$

$$\text{or } \phi = \frac{2\pi}{\lambda} n\lambda \sin\theta$$

$$= 2\pi\, n \sin\theta$$

So far we have considered only the aerial, but we must now allow for the effect of the image. As explained previously, we can take into account the effect of the earth by assuming an image aerial carrying a current in phase opposition to that in the actual aerial when we wish to obtain an approximate diagram of low angle radiation and one carrying a current in phase for high angle radiation. The in-phase case merely requires that the length of the aerial be taken as twice its actual length.

If the image oscillates in opposite phase, adjacent vectors at O (the earth) are exactly in anti-phase and a similar but reverse arc of vectors for the image is obtained as shown in Fig. 96c, the resultant vector being $A'B'$.

Draw $O'X$ tangential to $O'A'$ to cut the line $A'B'$. It will bisect $A'B'$ at right angles and $O'X$ will also be at right angles to $O'N$ since $O'N$ is the radius of the circle and $O'X$ the tangent.

Since angle $NO'X$ is a right angle and the angle $NO'A$ is equivalent to $\pi/2 - \phi/2$, the angle $A'O'X$ equals $\frac{\phi}{2}$. Now

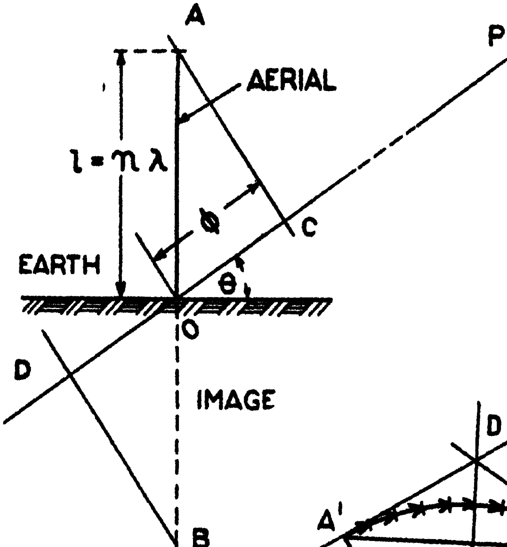


FIGURE 96a.

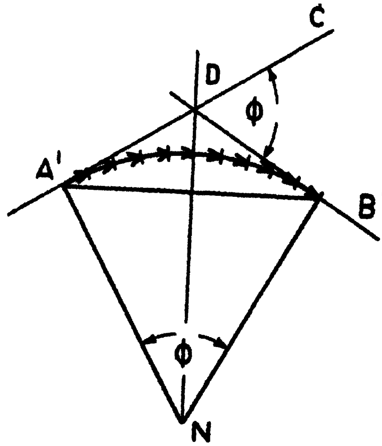


FIGURE 96b.

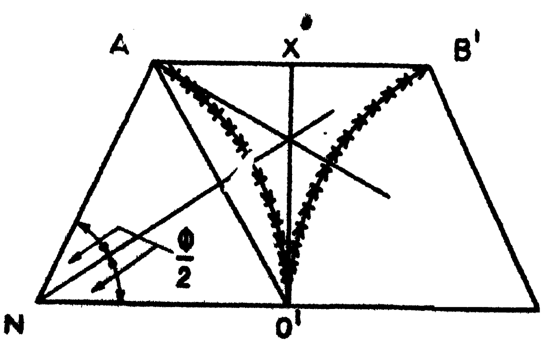


FIGURE 96c.

$A'X = A'O' \sin \phi/2$, therefore $A'B' = 2A'O' \sin \frac{\phi}{2}$. But $A'O'$ is the chord and equals $\dfrac{\sin \frac{\phi}{2}}{\frac{\phi}{2}}$ (since the arc is assumed of unit length).

Thus
$$A'B' = \frac{2 \sin \frac{\phi}{2}}{\frac{\phi}{2}} . \sin \frac{\phi}{2}$$

$$= \frac{2 \sin^2 \frac{\phi}{2}}{\frac{\phi}{2}}$$

The maximum possible field strength at P is, however, twice that due to the aerial alone, owing to the image. Also we have not, so far, allowed for the fact that the radiation from each element of the aerial will follow a cosine law in the vertical plane. Allowing for these facts we have

$$\frac{\text{Actual field strength at } P}{\text{Max. possible field strength}} = \frac{\sin^2 \frac{\phi}{2}}{\frac{\phi}{2}} \cos \theta$$

where the relationship between ϕ and θ has been shown to be $\phi = 2\pi\, n \sin \theta$.

As it is usually more convenient to work in degrees than radians, the above formula should be modified thus:

$$\frac{\text{Actual field strength at } P}{\text{Max. possible field strength}} = \left\{ \frac{\sin^2 \frac{\phi}{2}}{\frac{\phi}{2}} . \cos \theta \right\} \times 57 \cdot 3$$

where ϕ and θ are measured in degrees.

We have already noticed that an aerial cannot carry a pure stationary wave, since there is energy radiated along its length. When several half-wave aerials are in series, as in the tiered aerial, then there must be a considerable travelling

wave entering the aerial at its foot in order to supply energy to all the sections, so that at the foot of the aerial the travelling wave is a considerable component of the resultant, whilst near the top it is negligible.

In consequence of this the maximum radiation from a tiered aerial is not perpendicular to its length, even if earth effects are neglected, and the actual tilt of the main lobe of the polar diagram is dependent upon the frequency supplied to the aerial.

The effect is specially marked in the Franklin uniform aerial where, by altering the applied frequency (or, alternatively, changing somewhat the lengths of the various portions of the aerial), the main radiation can be varied in direction considerably. Some experimental results indicated that a 6% wavelength change could swing the angle nearly 10°.

Horizontal Aerials. So far we have considered only vertical aerials, but horizontal aerials are also greatly used. They are usually fed through an unearthed feeder of the parallel wire type, either designed for high efficiency at a spot wavelength, or for moderate efficiency over a band of wavelengths, and a variety of simple arrangements is possible.

Thus Fig. 97a shows a horizontal half-wave aerial whose actual length l, as explained on page 169, will be less than $\frac{1}{2}\lambda$, usually $\cdot 47\lambda$ to $\cdot 475\lambda$, fed from a twin feeder connected to points XY on the aerial such that a correct feeder termination is obtained. The correct distance XY depends on the characteristic resistance of the feeder used and on the height of the aerial above earth. If a 600 ohm feeder is employed and the distance of the aerial above earth is $\frac{1}{2}\lambda$ or more, $XY = \cdot 125\lambda$, whereas if the aerial is nearer earth, say $\frac{1}{4}\lambda$, $XY = \cdot 1\lambda$ for the same feeder. It should be observed that the depth of the bight between the aerial and the commencement of the parallel wire feeder proper has an influence on the tapping position XY and a length of bight $\cdot 15\lambda$ will be found correct for the dimensions previously given.

In the case of a transmitting aerial the feeder wires are run parallel (two 14-gauge wires separated 6″ giving 600 ohms is a common dimension), but with feeders for receiving work it is most desirable to use a transposed feeder, either by transposing the individual wires about every 2 metres, or by using a twisted flexible wire. An equally important precaution is

to make sure that the wires run are exactly the same length, both these measures being desirable in order to prevent unwanted "pick-up" on the feeder.

Such an aerial is highly efficient at wavelengths near the natural tune of the horizontal wire, but the efficiency falls away rapidly on either side of the tune. Since the feeder is correctly terminated its length is not important.

An alternative "general purpose" arrangement, usually only used for receivers, is shown in Fig. 97b. In this case the overall length of horizontal aerial, that is $2l_1 + l_3$, will be made approximately equal to one-half of the most important wavelength within the band it is desired to cover, and the dimen-

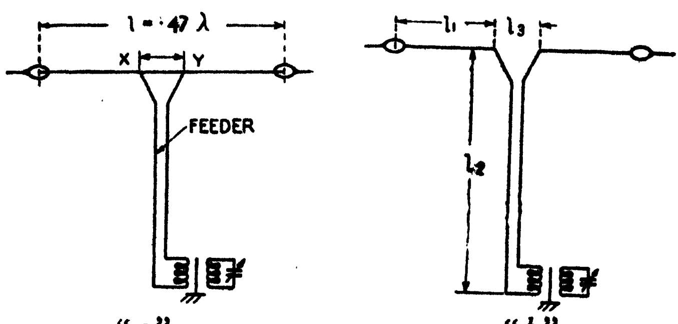


FIGURE 97.

sions of such an arrangement will be as follows: $2\,l_1 + l_3 = \frac{1}{2}\lambda$; the connecting piece inserted at $l_3 = .3\lambda$; and the length of feeder l_2 must not be longer than necessary, since it is for the most part incorrectly terminated, but it should be dimensioned such that it is ·5λ or 1·0λ if tapped into a parallel resonant circuit at the base; or ·25λ, or ·75λ if tapped into a series circuit at the base.

Such a "general purpose" aerial may be regarded as operating as arrangement (*a*) on and near the optimum wavelength, but at longer waves the feeder is not terminated, but now assists in the reception and in consequence the length of the feeder will determine the performance of the aerial as a whole. Thus we can regard the aerial as made up of two parallel aerials of length l_1 and l_2, whose various natural resonance wavelengths can be found as previously

indicated, and knowing these we can determine the type of circuit required to match the aerial. Because there are two down leads, vertically polarised interference will be eliminated, and in consequence such an aerial will give a somewhat better signal/noise ratio than a simple vertical wire.

In cases where a simple directive aerial is desired, it is possible to use arrangement 97b, but increase the length of each limb of the aerial to $\frac{1}{2}\lambda$. In this case the end of feeder will not be fanned out but run parallel right up to the aerial connections, which are each now end fed (see page 192). Since the feeder line is now not correctly terminated even at the optimum wavelength of the aerial system, the length of feeder must be designed correctly to fit the type of terminating circuit at the base, as indicated on page 188.

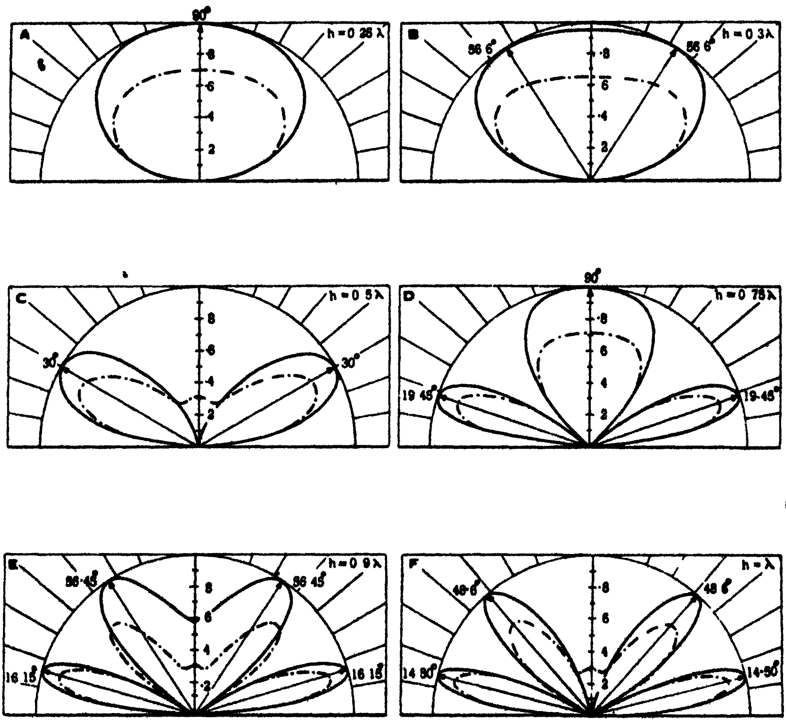
Radiation from Horizontal Aerials. If the effect of the earth is neglected altogether, the polar diagram of a horizontal dipole is the same as for a vertical one, but turned through a right angle so that the zenithal polar diagram becomes the horizontal diagram and vice versa. The polar diagram in a horizontal plane through the dipole is therefore a "figure of eight" and the dipole is seen to be directional.

We have already noted that the electric and magnetic fields radiated from a dipole are perpendicular to each other and both perpendicular to the direction of propagation. It follows that a vertical dipole radiates vertically polarised waves in all horizontal directions but the radiation in directions making an angle with the horizontal contains both vertically and horizontally polarised components.

If the dipole is now turned through a right angle it will be seen that the horizontal radiation in a direction normal to the axis of the dipole will be horizontally polarised but that the horizontal radiation in other directions will contain both horizontal and vertical components.

Let us now consider the effect of placing a horizontal dipole a short distance above a perfectly conducting earth. The propagation of horizontally polarised waves along the surface of such an earth would not be possible because it would involve differences of potential over a perfect conductor since the electric field is horizontal. Eddy currents would be produced, the field from which would exactly cancel out the original

field. For usual values of the earth's conductivity the horizontal radiation is very small and the zenithal polar diagram for a given aerial position relative to earth is less dependent upon the earth's constants than in the case of a vertical aerial.



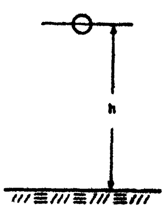


FIGURE 98.—EARTH CONSTANTS.

——— Perfect Conductivity $\sigma = \infty$

— · — · — Average Earth $\begin{cases} \sigma = 10^7 \text{ E.S.U.} \\ \kappa = 5 \text{ E.S.U.} \end{cases}$

Curves are deduced assuming mathematical dipole at height h above the earth as indicated.

A	$h = \cdot 25\lambda$	B	$h = \cdot 30\lambda$
C	$h = \cdot 50\lambda$	D	$h = \cdot 75\lambda$
E	$h = \cdot 90\lambda$	F	$h = 1\cdot 0\lambda$

The polar diagram at distances large compared with the height above the earth may be found by the image method previously discussed for vertical aerials, but in the case of the horizontal aerial we must reverse the sense of image current

as against that used for a vertical aerial over a similar earth, as has been explained in Chapter IV.

Figure 98 shows a series of zenithal polar curves calculated for a "mathematical dipole" at different heights above a perfect and imperfect earth whose values of σ and κ are as indicated in the figure. From these curves it is very evident that as one is chiefly interested in low angle radiation there is considerable advantage in raising an aerial a half wavelength above earth and that very little additional advantage will be gained by raising it above; unless the aerial can be raised to a height of at least one wavelength or more. Aerials raised between a half and one wavelength above earth have, as can be seen, strong radiating properties at high zenithal angles. An interesting feature is that neither the earth conductivity nor the wavelength have a vast influence on the shape of the polar curves. Although these curves are calculated for an idealised dipole they are substantially true for the ordinary half wave type of horizontal aerial.

It is known that whether waves are radiated from horizontal or vertical aerials, they are usually circularly polarised when they leave the ionosphere, hence, at distances beyond the skip area, either type of aerial may be used for reception from a transmitting aerial of either type. One type of fading is produced by changing polarisation and can be partially overcome by summing up the E.M.F.'s received on two aerials, one vertical and one horizontal, for tests made show that interference fading on a horizontal aerial is often opposite to that on a vertical aerial near it, and receiving from the same station. Generally a horizontal aerial provides a better signal/noise ratio than a vertical because certain electrical apparatus, such as ignition systems or sparking commutators on machines, produce highly damped E.M. waves of high frequency, which appear to be vertically polarised.

The Zeppelin Aerial. A form of feeder-aerial combination popular with amateurs but not much used by commercial organisations is as shown in Fig. 99, termed the Zeppelin aerial. It will be seen to consist of an aerial one half wavelength long (or a multiple thereof) and a parallel wire feeder, one limb of which is connected to one end of the aerial.

Since the aerial is fed at a voltage antinode, a feeder must be provided of the correct length to produce the voltage antinode at its far end; it is clear that a feeder of any length may be used, although it is highly desirable to keep it as short as possible consistent with the height of aerial, provided the ground or feed end be terminated by the correct tuning system to produce stationary waves as required.

If the curve "O," Fig. 68, page 137, be referred to, we can determine the type of circuit necessary at the feed end. For instance if a feeder just less than one quarter wavelength is used, the feed end will be capacitative and of small value,

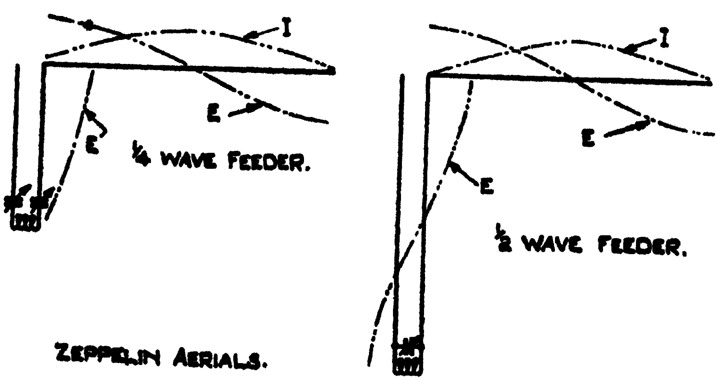


FIGURE 99.

and a small coil coupling will suffice. Actually, series tuning condensers can be provided so that this can be adjusted to give maximum current in the feeder at this end. If a feeder nearly one half-wavelength long be used, since the feed end is now of high impedance, a parallel feed circuit will be required, the current in the feeder now being very small for optimum adjustments.

Thus in setting up a feeder and aerial of this type, having a knowledge of the length of the feeder in terms of the wavelength, we can decide easily upon the type of feed circuit to be used. It should be observed that a feeder system operated under such conditions will tend to radiate, but the radiation can be minimised by balancing the current in the two wires to exact equality. This point has previously been mentioned.

Single Wire Feeder. It is possible to feed a half-wave aerial through a single wire feeder and so adjust the position of the tap that the current in the feed wire is free from stationary waves. Consider a half-wave aerial as shown in Fig. 100, on which is shown stationary waves of current and voltage. Consider the impedance between any point *A* on the aerial and earth. We can regard two circuits as connected to the point *A*. An open-circuited wire *AB* (rather less than quarter wavelength in the case shown); and a second open-circuited wire *AC* (the same amount greater than one quarter wavelength). If now the reactance curves on page 171 are consulted, it will be observed that wherever the point *A* be chosen the reactance of the length *AB* is always equal and opposite to that of the length *AC*, referred to the point *A*. If the tap is near the centre, these reactances are small and as the point *A* is moved away from the centre, the reactances rise to large values, but at all points the effective reactance is zero, and thus we are always "looking into" a resistive circuit wherever the point *A* be placed. The values of this effective resistance are low at the centre and rise to the high values associated with the base resistance of a half-wave aerial at the end.

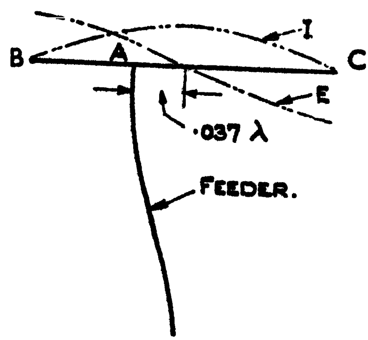


FIGURE 100.

In feeding an aerial through a single wire feeder, provided the feeder is tapped at such a point along the aerial that it is "looking into" a resistance equal to its own surge resistance, the feeder termination is correct and no stationary waves are present in the feed wire.

As shown in a previous section a vertical wire of normal dimensions near earth has a characteristic resistance of about 500 ohms and thus the point "*A*" will be chosen to match this. A point about ·037λ along the aerial either side of the centre will be found to be nearly correct.

It is evident that there will be radiation from the single wire feeder, but if the tapping point has been correctly adjusted

to eliminate stationary waves, the current in the feeder will be much less than in the centre of the aerial and hence the greater part of the radiation takes place from the aerial, as desired. The adjustment of the tapping point to eliminate stationary waves is very critical and it will be realised that mal-adjustment is more serious than with the Zeppelin aerial because of the radiation from the feeder that results.

Aerial Matching by a Reactive Stub. We have seen in a previous chapter that, where possible, it is highly desirable to match feeder systems to output circuits, not only to avoid increased losses in the system but also to prevent radiation

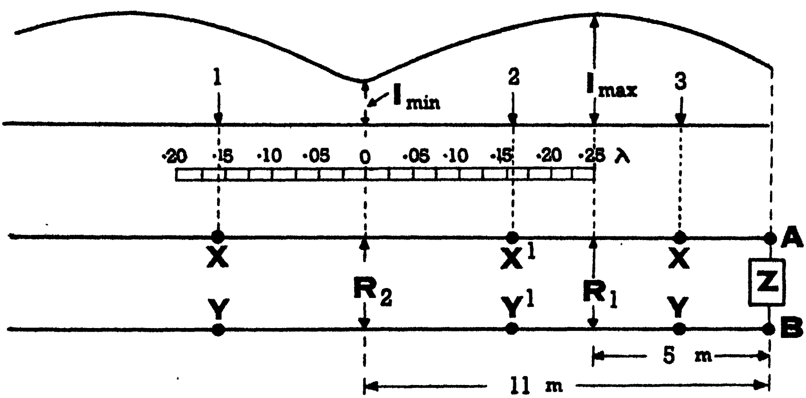


FIGURE 101.

from the feeder itself. This can be done either by using matching circuits of the type described in Chapter VI or by the use of what are called "matching stubs."

From Chapter VI it was made clear that the input impedance of a length of short-circuited (or open-circuited) feeder will have a pure reactance of value dependent upon its length. It is, therefore, possible by means of such a stub feeder length to balance out the reactance of a mis-terminated feeder by placing the stub at a suitable point along the feeder, preferably as close to the termination as possible.

Consider a feeder of characteristic impedance R_o terminated at AB by a complex impedance Z of $R + jX$ as shown in Fig. 101. Since the termination is incorrect, a stationary wave system will be present which may take the form, say, as indicated.

If we examine the conditions along the line from the termination towards the input we could consider the problem as if we had a feeder of a different length terminated by pure resistance at those points where the stationary wave current values are a maximum and a minimum, the equivalent pure resistance value R_1 at the current maxima points being lower in value than R_0 and the resistance value R_2 at current minima points being higher than R_0. The amount by which the load impedance departs from the correct terminating value is shown up by the extent to which the ratio of currents I_{max} and I_{min} differs from unity. Since at the current-maximum point the equivalent resistance R_1 is lower than R_0 and at the current-

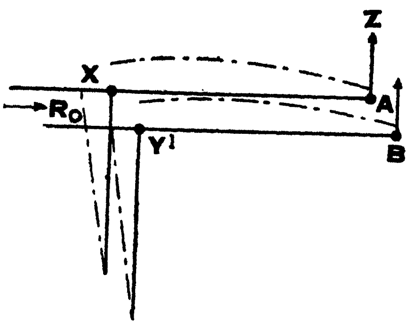


Figure 102a.

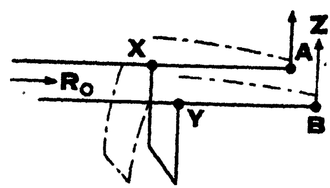


Figure 102b.

minimum point higher, it is clear that at some point XY or $X'Y'$ between, we are looking into an impedance which consists of a resistance of R_0 together with a reactance of some unknown value. If one of these points can be found, then by placing across the line at that point a pure reactance of the same value as the unknown but of opposite sign, we shall obtain a network such that the feeder will now be "looking into" a circuit of equivalent resistance R_0 and stationary waves will be eliminated on the line from the input up to this point.

Matching circuits using stub lines are shown in Figs. 102a and 102b, where the length of stub connected across the points XY (or $X'Y'$) is arranged to cancel the reactance of the aerial, together with the length of feeder between the stub and aerial, so that the main feeder looks into a pure resistance of R_0, equal to its characteristic impedance.

The choice of stub, that is, whether of the shorted or open-circuited variety, is dependent to some extent upon the position of the stub relative to the terminal and the length of the stub itself. For most cases the shorted stub is more desirable than the open-circuited, because it is easier to adjust, and the centre of it, which is at low potential, can be anchored to earth if necessary or supported by a mechanical structure which does not affect its electrical properties. If, however, it is found that the length of short-circuited stub required for balancing requirements is less than about ·1 of a wavelength, it will not be used and the open-circuited stub is to be preferred. This is because a very short stub can hardly be regarded as a transmission line system. When installing open-circuited stubs, however, it is found that the holding insulators have a considerable influence on the stub length and because open-circuited units are at high voltage, insulator break-downs are liable to occur with large powers unless adequate precautions are taken.

The correct position of the stub and its length can be obtained theoretically from the equations developed in Chapter VI, but a graphical solution has been given in a series of curves first derived by Sterba and Feldman. A simpler modification of these is given in Fig. 103.

With these curves it is only necessary to measure the relative values of I_{min} and I_{max} and their positions from the load in order to determine directly from the curves: firstly, the position of either a short-circuited or open-circuited stub in relation to a current minimum; secondly, the length of stub required.

In practice it is always desirable to work from a current-minimum position as can be seen by studying the shape of the stationary wave system along the line where it is observed that the rate of change of current about a minimum point is much sharper than about a maximum point.

The method of using the curves can best be shown by an example.

EXAMPLE.—Consider a length of feeder terminated by an impedance Z such that a stationary-wave system is set up as shown in Fig. 101. That is, measurement by means of a sliding meter indicates that there is a current-maximum point

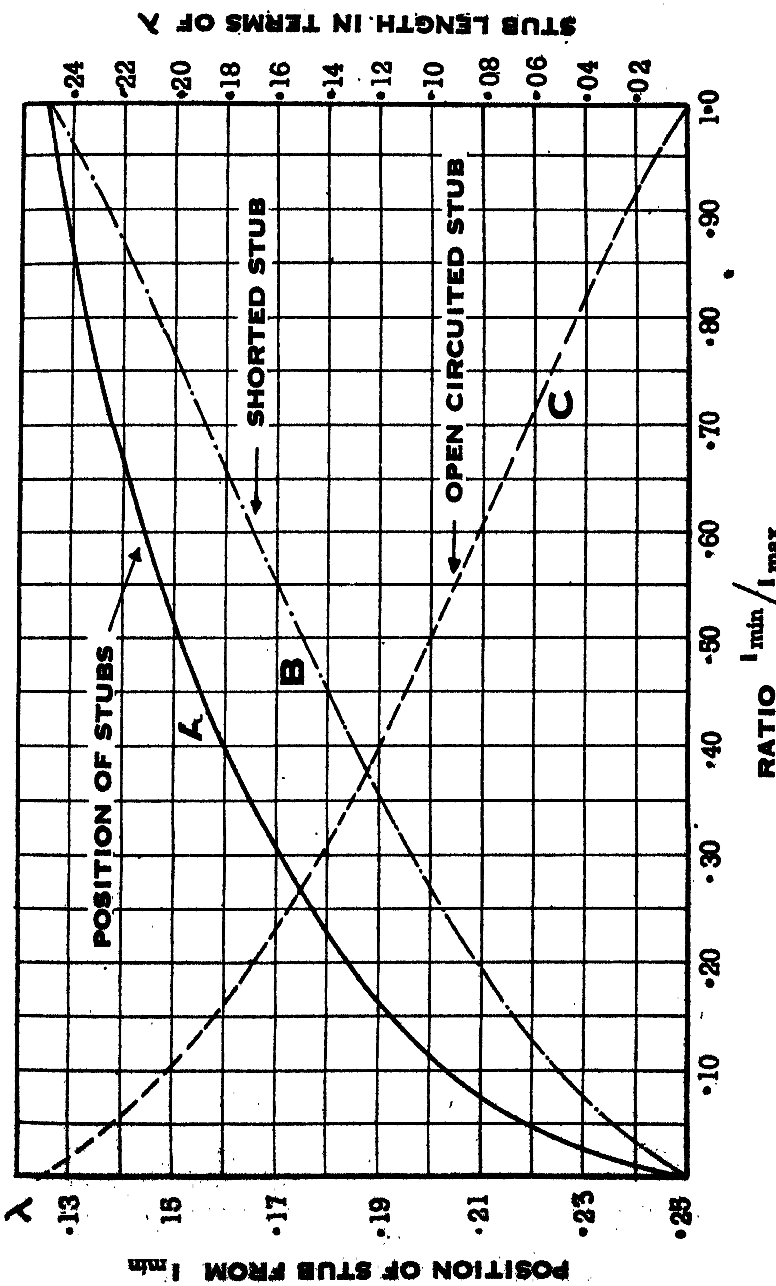

STUB LENGTH IN TERMS OF λ
·24
·22
·20
·18
·16
·14
·12
·10
·08
·06
·04
·02
SHORTED STUB
OPEN CIRCUITED STUB
POSITION OF STUBS
A
B
C
λ
·13
·15
·17
·19
·21
·23
·25
POSITION OF STUB FROM I min
·10
·20
·30
·40
·50
·60
·70
·80
·90
1·0
RATIO I min/I max

FIGURE 103.

5 metres from the load, of value 1·0 *A* and an adjacent current-minimum point 11 metres from the load of value $= 0{\cdot}4$ A, the wavelength being 24 metres, the ratio $\frac{I_{min}}{I_{max}} = 0{\cdot}4$.

From curve *A* Fig. 103 it is found that with this ratio the stub length must be placed at a distance of 0·16λ from a current minimum, i.e. either at position 1 or 2.

If a short-circuited stub is selected, this distance must be measured away from the load towards the input, whilst if the stub is to be open-circuited, the distance will be measured towards the load.

From curve *B* we find the length of a shorted stub to be 0·13λ and it could be placed at position 1. In this example, however, 1 is more than $\frac{\lambda}{2}$ from the load and the stub will be equally effective if placed at 3—which is $\frac{\lambda}{2}$ from 1. Position 3 is evidently the better, because there is then a shorter length of feeder carrying a stationary wave.

Curve *C* shows that the length of an open-circuited stub will be 0·12λ, and we know this must be placed away from the input, namely at position 2.

Coupling to an Unbalanced Aerial Load. Although normally such a terminating circuit as has been described will be used for a parallel wire feed to a balanced aerial system, such for instance as a stacked di-pole array, it is possible to adapt it for use with an unbalanced vertical aerial taken direct from one side of the feeder, Fig. 104 giving one possible form of circuit. The length of wire *AB*, one end of which is connected to *A*, the base of the aerial, is half a wavelength long, so that it reflects an equal reactance at the point *B*, but in opposite phase, and thus the feeder may be regarded as looking into a balanced complex impedance. There will, of course, be stationary waves on this wire length as well as on the stub and feeder end.

Some Points in Design and Use of Aerials. Enough has been said to show that aerials should be designed for the wavelength at which it is desired to work, but this is, however, not always possible and efficiency may have to be sacrificed to practical convenience. A marine transmitter or receiver,

for example, will have to work over a large range of wavelengths and we cannot rig up a special aerial for every wave and may even have to work on the ship's long-wave aerial. For reception, the aerial may be coupled to the receiver through a length of low-capacity cable, the aerial circuit being untuned. The type of horizontal aerial described in a previous section is also suitable.

For transmission, the ship's main aerial may be used as a harmonic aerial, coupling its lower end through a series or parallel circuit as described in the first section of the chapter. If we know the length of aerial available, we can very quickly see at what wavelengths such an aerial will be efficient, and possibly by a small alteration of length, which will not effect

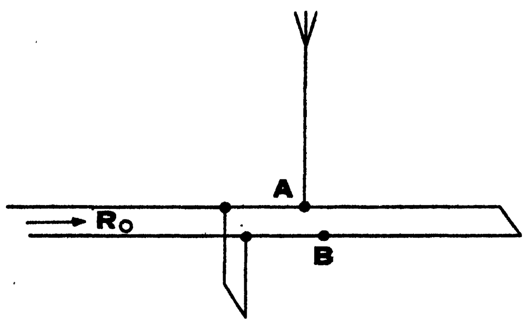


FIGURE 104.

its working on long waves, efficiency can be obtained at the particular short waves desired.

It will be appreciated that long "leading in" wires, or "trunks," which are a common feature of the long wave system, may be most inefficient on short waves, as the capacity of these leading in arrangements may be sufficient to bypass most of the current. Their length may be favourable to the formation of a large stationary wave due to reflection at the deck insulator, where the inductance and capacity per unit length change greatly in value. Where possible it is always better to use a feeder, commencing the aerial proper where the surroundings are clear, although the use of such will necessitate a terminating circuit at the bottom of the aerial. If this cannot be done, correct termination at the transmitter end of the trunk may enable more power to be fed to the aerial.

In short wave work it is not necessary to take great precautions to keep down the resistance and earth losses, so necessary in the case of the long wave, because the radiation resistance is so much greater. Hence arrangements such as earth screens will not be used.

SELECTED REFERENCES

Short Wave Wireless Telegraphy. T. L. Eckersley. *J.I.E.E.*, Vol. 65 ; June, 1927.

The Radiation Resistance and Energy Capacity of Half-Wave Aerials. E. Green. *W.E.*, Vol. 5 ; February, 1928.

Distribution of Current in a Transmitting Antenna. Wilmotte. *J.I.E.E.*, Vol. 66 ; September, 1928.

The Nature of the Field in the Neighbourhood of an Antenna. Wilmotte. *J.I.E.E.*, Vol. 66 ; September, 1928.

Radiation from a Short Wave Vertical Aerial. T. L. Eckersley. *Mar. Rev.*, No. 23 ; August, 1930.

Graphical Method of Determining Magnitude and Phase of Electric Field in Neighbourhood of Antenna. McPetrie. *J.I.E.E.*, Vol. 69 ; February, 1931.

CHAPTER VIII

AERIAL ARRAYS

The term "aerial array" has been given to combinations of aerials arranged to produce some special directional effect. The whole array may in some cases be used simultaneously for the transmission (or reception) of more than one signal.

The advantages of directional transmission and reception have, of course, been realised from the early days of wireless, but efficient systems have only become possible since short waves came into use, as it is only possible to produce any efficient transmitting or receiving directional aerial-system when the size of the array is large compared with the wave-length being used.

General Requirements. When a long distance point-to-point service is under consideration, the horizontal polar diagram at the transmitter should be narrow, and the ratio of forward to back radiation should be as great as possible to concentrate the available energy into a narrow beam. The vertical diagram also should be sharp and it may be desirable to be able to change the directive properties in the zenithal plane so as to be able to select the particular ray predominant at any one time.

At both transmitter and receiver there is a limit to the narrowness of the horizontal diagram, because we have evidence that the apparent direction of the transmitting station does not always remain quite constant.

Since the arrays are probably the most expensive part of a short wave station, it is very desirable in the interests of economy to be able to couple a number of receivers to the same array, or to be able to use a transmitting array on more than one frequency. This means that the polar diagram should not be dependent upon precise tuning of any part of the array.

The direction of maximum horizontal radiation should not be too rigidly dependent upon the direction in space of the array. For it is evidently an advantage if, by a simple electrical adjustment, an array can be biassed to radiate in a different direction.

Some special cases may, of course, have peculiar requirements. An interesting example is the type of array required for the telephone service from Rugby to Atlantic liners. In this case the horizontal diagrams must evidently be wide enough to cover the area of ocean over which it is desired to maintain communication, and in the case of the longest wave array, which is used for the shorter distances, the beam has a large spread and is arranged to radiate principally at fairly high angles, as it is the high angle radiation which returns from the ionosphere in the shortest distance.

The most obvious way of forming a beam is to use a parabolic reflector with a single aerial at the focus.

This method of producing a beam, which was the first employed, is now used only for ultra-short waves, as newer systems have been developed which possess many advantages, chief of which is the simpler mechanical construction, and the greater gain of field strength produced by the uniform division of energy among a number of aerials.

Array Systems. The modern aerial array system, of which C. S. Franklin was the originator, obtains its " beam " effect by the grouping of a number of radiating elements fed by current in such phase as to produce an interference pattern in space giving the required directive properties.

In general, array systems are of two main types, the " Broadside " type and the " In-line " or " End-Fire " type. In both cases the array consists of a long line of radiating elements, but whereas the Broadside type concentrates the energy normal to the line of radiators, the latter is most active along the aerial line.

Before we consider the various systems in use, it will be of interest to illustrate the interference pattern produced by two spaced omnidirectional aerials.

Polar Diagram of Two Spaced Aerials. Consider the field produced at P (Fig. 105) by two aerials A and B (assumed to be point sources of radiation) spaced $n\lambda$ metres apart,

where λ is the wavelength in space corresponding to the frequency of the currents in the aerials, it being assumed that P is so far away that lines joining A and B to P are parallel. Evidently, if the currents in the aerials are in the same phase, the fields due to each aerial will not be in phase at P, because of the difference in distance, from P to each aerial.

This difference in distance is most conveniently expressed by the phase angle it produces between the fields at P due to the two aerials, and this phase angle ϕ will be termed the "space-phase."

It will be seen that the difference between the lengths PA and PB is $n\lambda \sin \theta$, and, since a difference in length of λ

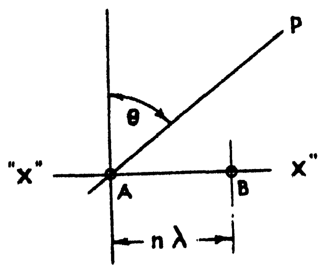


FIGURE 105.

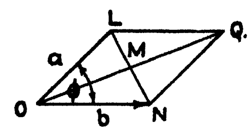


FIGURE 106.

would cause a phase angle of 360°, the phase angle ϕ actually produced will be $\frac{360^\circ}{\lambda} n \lambda \sin \theta$, or $\phi = 360^\circ \, n \sin \theta$.

If the currents in A and B are in phase and produce fields a and b at P which are equal numerically, then the vector sum at P is OQ, as shown in Fig. 106.

$$\text{Now} \quad OQ = 2OM \text{ and}$$

$$OM = ON \cos \frac{\phi}{2}$$

$$\text{Hence} \quad OQ = 2b \cos \frac{\phi}{2}$$

Applying these formulæ to spacings of $\frac{\lambda}{8}$, $\frac{\lambda}{4}$, $\frac{\lambda}{2}$, and λ, we get polar diagrams shown in Fig. 107, line one.

In many cases, however, the two aerials are supplied with currents which are out of phase in time so that there is a *time-phase* of α degrees. For such cases the resulting phase angle ϕ', say, between the two fields will be dependent upon

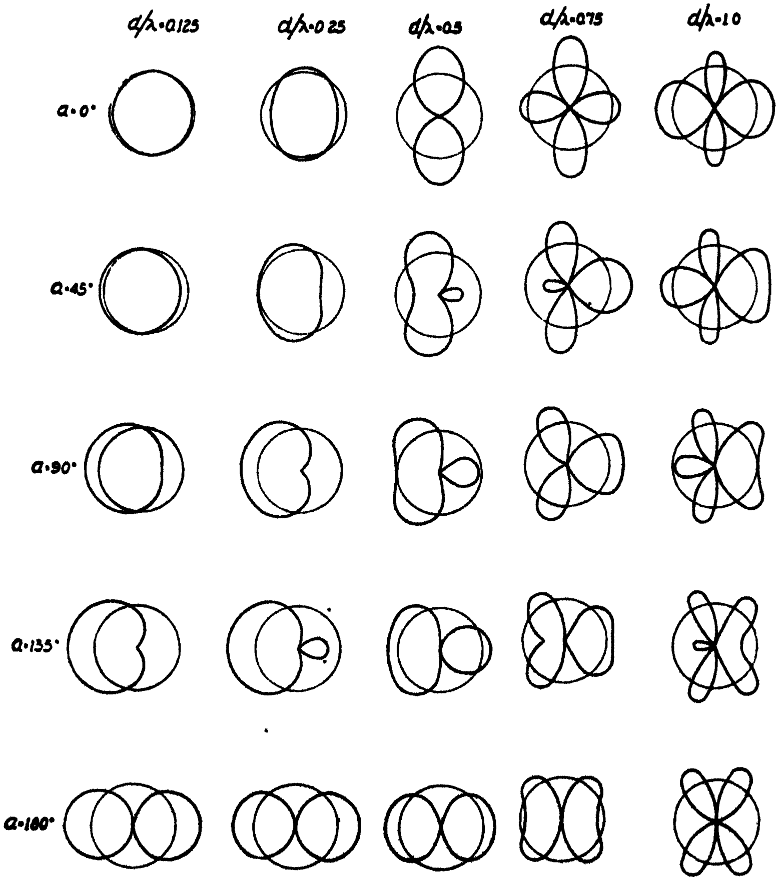


FIGURE 107.

whether the time phase α is additive to or must be subtracted from the space phase ϕ. For directions to one side of the normal the two will add, whereas on the other side they will subtract, that is,

$$\phi' = \alpha \pm 2\pi\, n \sin\theta \text{ radians}$$
$$\text{or} \quad = \alpha \pm 360\, n \sin\theta \text{ degrees.}$$

If the current in the more distant aerial is lagging behind that in the nearer one, then evidently ϕ is the sum of the space-phase and the time-phase, whereas, if the current is leading, ϕ is the difference.

With different values of spacing and time phase, polar diagrams such as shown in Fig. 107 are produced, the circles showing polar diagrams of a single aerial supplied with the same power. Time phases between 180° and 360° would give similar diagrams but reversed in sense along the XX axis. From these curves it is observed that with zero and 180° of time phase, diagrams bi-symmetrical both about the line of spacing XX and the line normal to it are produced. Further, with zero time phase maxima values are always produced in a direction normal to the line XX, whereas with 180° time phase, minima values are produced always normal to the line of aerials XX. For phases in between 0° and 180° (and 180° and 360°) diagrams asymmetrical to the normal line are produced, such diagrams having optimum uni-directional properties along the line XX when the time phase is 90° or 270°.

The Inter-Action of Two Adjacent Aerials. The polar diagrams obtained above are correct if the currents in the two aerials have actually the relative phases specified, but it is necessary to realise that the two aerials are coupled together, due to each being in the field of the other, this fact being allowed for in computing the diagram areas.

Thus, if we first tune each aerial in the absence of the other and then apply to both of them E.M.Fs. in the same phase, each aerial will induce an E.M.F. in the other, the phase of which will depend upon the spacing. The currents flowing will be due to the resultant of applied and induced E.M.Fs.

The problem can be worked out as a coupled circuit, introducing the notion of a mutual impedance Z_{12} between aerials 1 and 2. The calculation of Z_{12} is a complex matter, involving, as it does, integrating the effect of the total field (and not merely the radiation component) produced by one aerial on each element of the other, but the values of Z_{12} for various aerials and spacings have been worked out. As an

example, Fig. 108 (derived from the paper of G. H. Brown) shows the resistive and reactive components of the mutual impedance between two $\lambda/2$ aerials, both at the same height

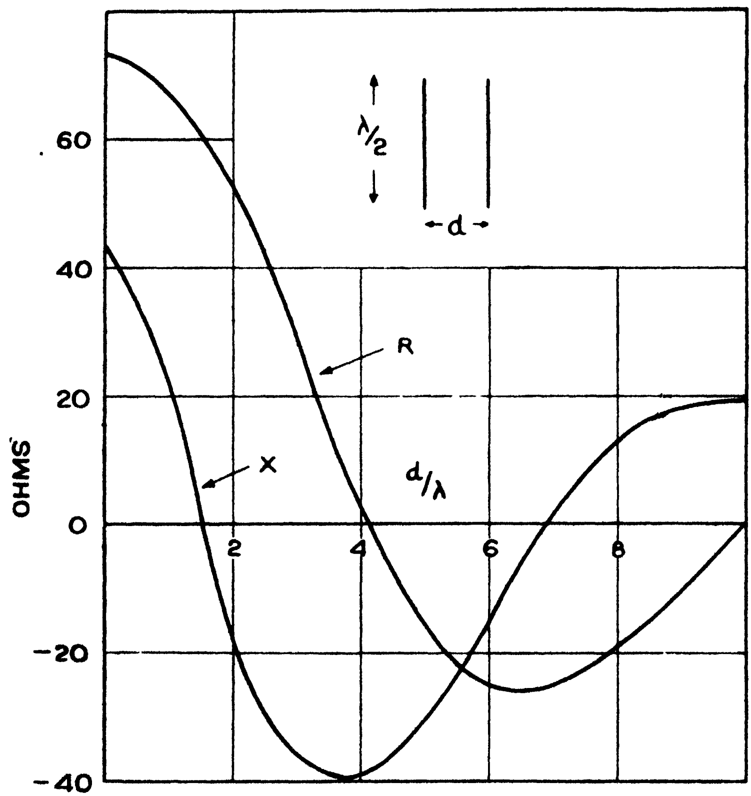


FIGURE 108.

and sufficiently elevated that the effect of the earth on the mutual impedance may be neglected.

Similar curves for other cases are given by Brown and by Carter. Supposing Z_{12} to be known, then as in other coupled circuits:

$$\left.\begin{aligned} E_1 &= I_1 Z_1 + I_2 Z_{12} \quad (1) \\ E_2 &= I_2 Z_2 + I_1 Z_{12} \quad (2) \end{aligned}\right\} \text{These are vector equations.}$$

If $E_1 = E_2$ in magnitude and phase and both aerials are independently tuned, so that $Z_1 = Z_2 = R$; then $I_1 = I_2$ in magnitude and phase, as would be the case if there were no interaction. It is now seen that the aerials when together are no longer in tune, in the sense that the impedance each presents to the circuit from which it draws its power is no longer a pure resistance. The currents are therefore different than for one aerial alone. The impedance is

$\frac{E_1}{I_1}$ and from (1) is $Z_1 + Z_{12}$ (since $I_1 = I_2$). Let $Z_{12} = R_{12} + jX_{12}$, then the aerials can be again brought into tune by altering their length so that Z_1 becomes $R - jX_{12}$. The aerials will now form a resistive load of value $R + R_{12}$ (R_{12} can be positive or negative, depending upon the spacing). Since the currents and effective resistances are not the same as for one aerial by itself, the power radiated is not twice that radiated by one aerial alone.

Aerial and "Reflector." It is frequently convenient to make use of reflector or parasitic aerials which are not supplied directly with power, but the currents flowing in them are caused entirely by the E.M.Fs. induced from neighbouring aerials. The relationships now become

$$E_1 = I_1 Z_1 + I_2 Z_{12} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$0 = I_2 Z_2 + I_1 Z_{12} \quad . \quad . \quad . \quad . \quad . \quad (4)$$

From which $$\frac{E_1}{I_1} = Z_1 - \frac{Z_{12}^2}{Z_2}$$

$$I_1 = \frac{E_1}{Z_1 - \frac{Z_{12}}{Z_2}} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$I_2 = -\frac{E_1 Z_{12}}{Z_1 Z_2 - Z_{12}^2} \quad . \quad . \quad . \quad . \quad (6)$$

Since these are vector equations, they give us the magnitude and phase of the currents and we can then construct the polar diagram by the method previously given.

A very wide variety of effects are evidently possible, since we can vary Z_{12} (by altering spacing) and both Z_1 and Z_2 (by altering aerial tuning—adjusting the length, for example).

The polar curve of an exactly-tuned $\frac{\lambda}{4}$ aerial and reflector spaced $\frac{\lambda}{4}$ apart has a shape between that for $a = 90°$ and $a = 135°$ in Fig. 107.

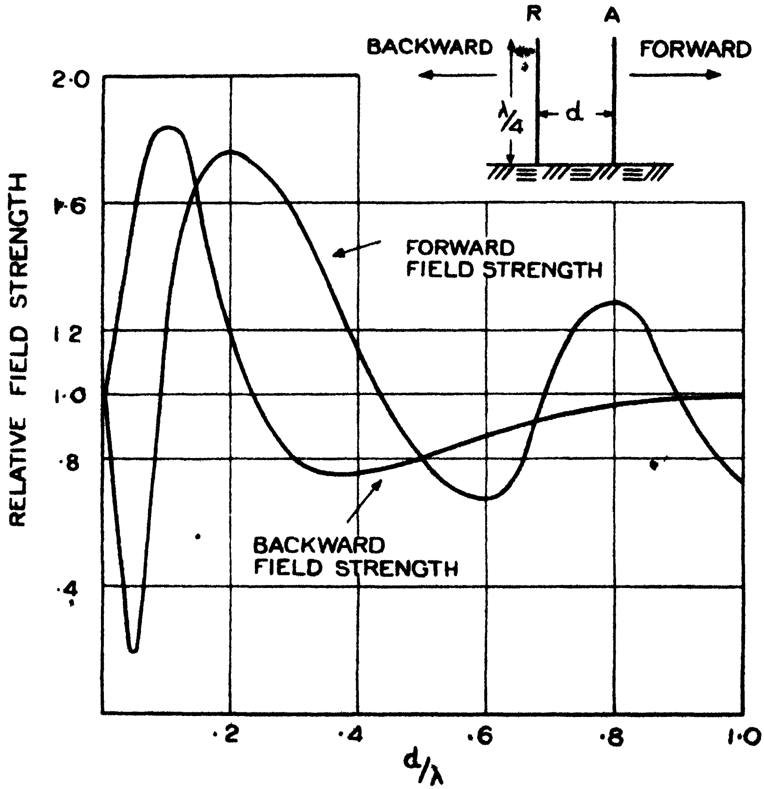


FIGURE 109.

If aerial and reflector are both resonant and the spacing is varied we obtain the curves of Fig. 109 for field strength in forward and backward directions. It will be seen that the parasitic aerial can act as a "director" as well as a "reflector" as the spacing is varied.

When the aerial and reflector are used for reception (the aerial being connected to the receiver) the directional

properties are similar to those for transmission, but the actual performance is evidently dependent upon the way in which the receiving circuit loads the aerial.

Polar Diagram of a Line of Radiators. The foregoing discussion relating to two single radiators illustrates the interference principle, but the diagrams obtained show equal energy concentration in various directions and are therefore not suitable for directional working. We can, however, eliminate the energy concentration in all except one main

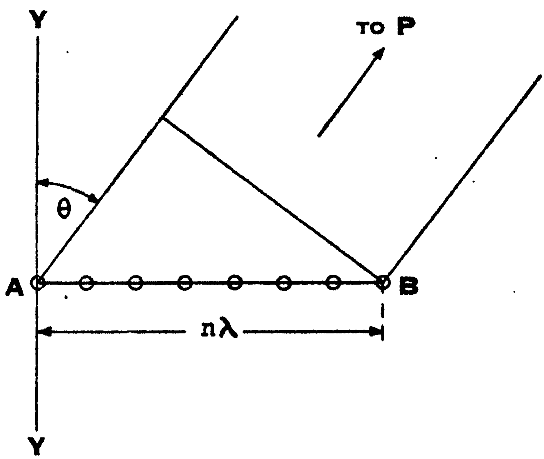


FIGURE 110.

direction by employing a number of radiators spaced along the line.

For instance, consider a line of eight aerials spaced each one quarter wavelength apart as shown in Fig. 110, these eight aerials making an array of effective length between the first and last aerial of $n\lambda = 2$ wavelengths. It will be found that such a line of closely-spaced aerials will concentrate the energy into one main beam and a number of subsidiary "tails" each side of the line, and, moreover, the main beam can be made to assume any angle with the line of the aerials depending upon the time-phase of currents with which the aerials are supplied.

The field at a distant point P at any angle θ to the normal will, of course, be the vector sum of a field due to each aerial.

If we consider the important case when all the aerials carry currents in phase with each other, i.e. zero time phase, then the maximum field as shown by the central vector (Fig. 111) will obviously be along a line YY, normal to the line of aerials. At directions other than normal, and making, say, any angle θ with it, to the right or to the left, the vector group will become wrapped up either in a clockwise or anti-clockwise direction as shown by Figs. 111a, b, c and d, right and left. Thus, for small angles of θ, a vector figure such as a will be produced. At greater deviations the vector sum produces a zero resultant such as shown in Fig. 111b (for the case of

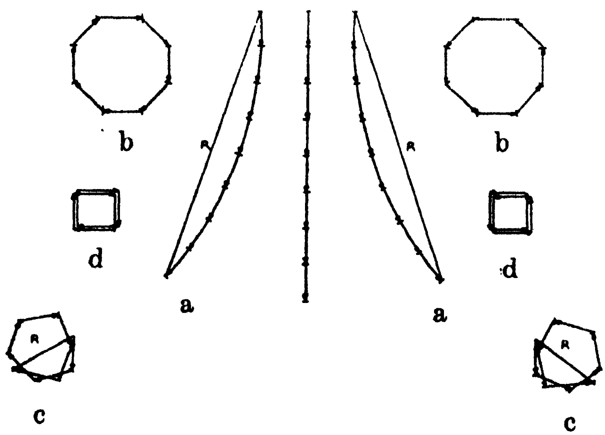


FIGURE 111.

$n\lambda = 2$, θ will be 30°). Thus the field has diminished from unity at $\theta = 0°$ to zero when $\theta = 30°$.

Considering still greater deviations from the normal, the vectors become still more wrapped up to produce vector figures as shown in Fig. 111c, where the resultant is equal to the diameter of a circumscribed circle, and 111d where the vector diagram becomes wrapped up twice giving a zero resultant, this occurring for $\theta = 90°$.

If we maintain the length of the array line the same but alter the number of aerials (and, therefore, the spacing), the vector diagrams are but little altered in shape. We can, therefore, suppose our actual aerials replaced by a very large number placed very close together, when the vector diagrams

will resolve into arcs and circles instead of polygons. This artifice, originally adopted by E. Green, has already been discussed on page 184 in connection with the zenithal polar diagram of a vertical aerial, and it will be evident that in the case of the array $n\lambda$ long the field strength in a direction making an angle $\theta°$ with the normal to the array line is related to the maximum field strength along the normal by

$$\frac{\text{Field at } P}{\text{Max field}} = \frac{\sin\frac{\phi}{2}}{\frac{\phi}{2}}$$

where $\phi = 2\pi n \sin\theta$ (radians)

or $360\, n \sin\theta$ (degrees)

Minima will occur where ϕ the phase angle between the field from the first and last aerial is 2π, or multiples thereof, namely at such angles that

$$\sin\theta = \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \frac{4}{n}, \text{ etc.}$$

and maxima will occur at angles such that

$$\sin\theta = 0, \frac{3}{2n}, \frac{5}{2n}, \frac{7}{2n}, \text{ etc.}$$

and the amplitude of the side loops relative to that of the main loop (unity amplitude) will be of value

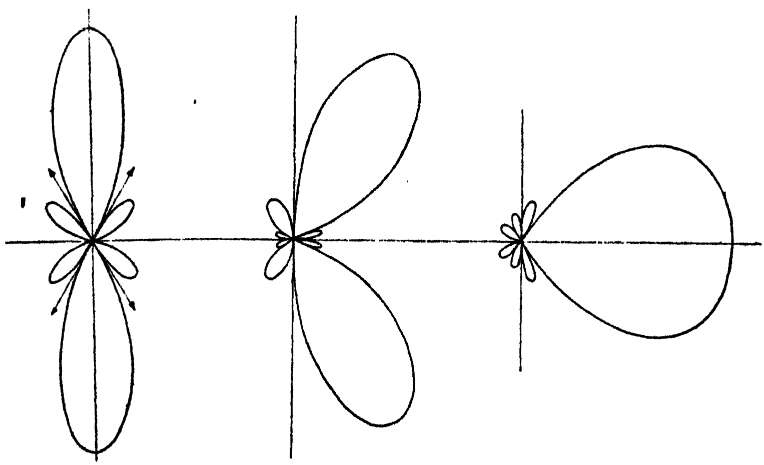
$$\frac{2}{3\pi}, \frac{2}{5\pi}, \frac{2}{7\pi}, \text{ etc.}$$

The polar curve for a two wavelength array of eight aerials each fed with current in time phase will be as shown in Fig. 112a, and it is observed that a perfectly symmetrical bi-directional polar diagram of radiation is produced, because the vectors wrap up either in a clockwise or an anti-clockwise direction from an initial straight-line vector both sides of the normal.

Let us now alter the time phases of the currents in the eight aerials of our typical array, so that the current in aerial 1 is leading by a small angle on that in aerial 2 and so on. This will result in the vectors having an initial bias because the

time phase reduces the field along the normal to the array line. (Fig. 111a, left, might now represent the conditions.)

Considering that direction (to the right) where the space phase between aerials is equal to the time phase we have applied, the time phase and space phase will be in opposite sense because in this direction the space phase of No. 1 vector is lagging on No. 2 and so on, whereas the time phase is leading. This will result in the vector sum being straightened again, thus giving maximum field for this direction. Contrariwise, at an equal angle θ from the normal in the left hand quadrants,



a *b* *c*

FIGURE 112.

the space phase is additive to the time phase and the field is reduced. The result of this small time phase is therefore to give a bias to the maxima of the polar diagram to that side away from the leading time end and make it asymmetrical as shown in Fig. 112b.

If we now increase the time phase to such an extent that it is exactly equal to the space phase between elements (in this case $\frac{1}{4}\lambda = 90$ degrees) the vectors for directions normal are the wrapped up vectors for Fig. 111c (left). Thus in directions to the right the vectors unwrap themselves and just become unwrapped for a direction in line with the array, as in this direction the space phase is now equal and opposite to the

time phase. To the left the vectors become more wrapped up still and produce small tails. Thus the polar diagram is biassed to such an extent that the maxima now overlap and produce a uni-directional polar curve lying along the array line shown by Fig. 112c, the maximum of this polar curve being away from the end which is leading in time phase, in this case No. 1 aerial.

If desired we can completely reverse this polar curve by changing the time phase such that the opposite end (No. 8) is leading in time phase, say by feeding from the end No. 8 instead of No. 1. Alternatively, if we still desire to feed the system from the No. 1 end, we must misphase each succeeding aerial by 360 degrees minus the space phase, as this is the same as giving the next aerial a lagging current.

We have considered the case of eight aerials each spaced $\frac{1}{4}\lambda$ apart. If we were to position our aerials more closely together we should obtain the same result if we altered the time phase to equal the new spacing. Thus if our aerials are arranged to be one-sixth of a wavelength apart, i.e. 60 degrees, we should need to produce a time phase between radiating elements of 60 degrees or (360 degrees + 60 degrees) to produce a maximum away from the fed end; or (360 degrees − 60 degrees) to produce a diagram with maximum towards the fed end.

As explained previously the directive properties are not materially changed with spacing of elements, and in general $\lambda/4$ spacing is customary.

Thus we have the general rule for a single line of radiators that with zero (and opposite time phase) or multiples thereof, bi-directional symmetrical figures are produced, the former with maxima normal, and the latter with maxima in line with the array.

If the time phase is made equal to the space phase between radiators a uni-directional polar curve is obtained with a maximum away from the feed or leading time-phase end. With a time-phase equal to 360 degrees minus the space-phase. the polar curve is reversed and the maximum now points towards the feed or leading time-phase end.

It is evident that there will be mutual impedances between each aerial and all the others and, if the aerials are to be in

tune, it will therefore be necessary to adjust the electrical length of the aerials to suit the spacing. This adjustment will be done experimentally, with previous experience as a guide, since the problem would evidently be very cumbersome to solve theoretically.

Broadside Arrays. It is clear that a single line of aerials, carrying currents in phase with each other, produce a bi-directional diagram. With such broadside arrays it is therefore necessary to provide a second line of radiation in the rear, to reinforce the forward radiation and cancel the backward.

It is usual (because more convenient) to supply only the front line and let the rear line act as " reflectors " or " parasites." Where we are concerned with lines of aerials it is found that a spacing of $\frac{\lambda}{4}$ is about the best to employ and the lengths of the reflectors are adjusted to give the best ratio of forward to backward radiation.

The tuning and spacing adjustments which would cancel backward radiation are not the same as those which would increase forward radiation to the maximum extent and a compromise is therefore adopted.

In Figs. 113 and 114 polar curves for a 2λ array, having reflectors $\frac{\lambda}{4}$ and $\frac{3\lambda}{4}$ respectively behind the active aerials, are shown. They assume that the reflector currents are equal in magnitude to those in the front aerials but are in quadrature leading for a $\frac{\lambda}{4}$ reflector ; and in quadrature lagging for a $\frac{3}{4}\lambda$ reflector.

From these figures it can be observed that the polar diagram is characterised by a main loop and a number of tails, and it will be remarked that the wider the array, the more concentrated is the main loop and the greater the number of tails. The width of an array system is termed its aperture, and this is defined in terms of the wavelength. Thus the wider the aperture, the better the main beam, but the more the number of subsidiary loops. An array built up vertically gives concentration of energy in the zenithal plane in exactly the same manner.

In assessing the value of an array we must be careful to take into account its current distribution. For instance, an

array of " A " square metres fed at one point, say the centre, would not be nearly so effective as the same area fed, say, at four points. For with a centre point feeder, because of radiation and general losses, there would be a tapering current to the edges, and in consequence the effective area would be reduced.

Array systems may be made to radiate either vertical or horizontally polarised waves and there are many examples of both types of arrays in practice.

Theoretically, as has been seen in a previous chapter, there should be but little to choose between the two types of waves

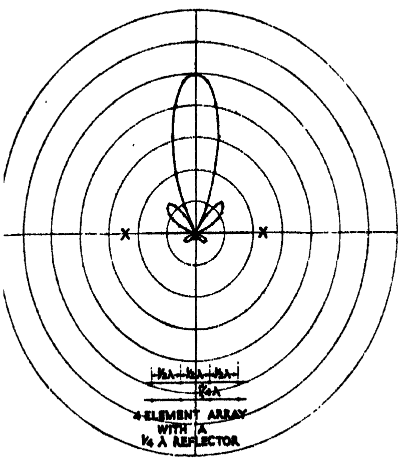


FIGURE 113.

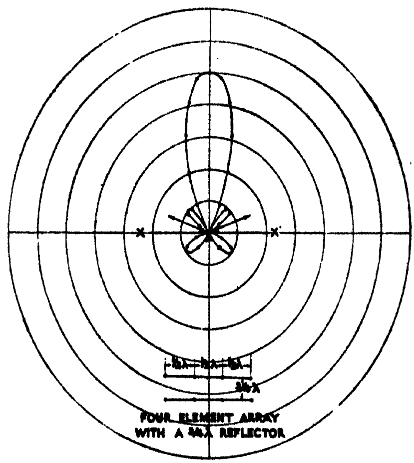


FIGURE 114.

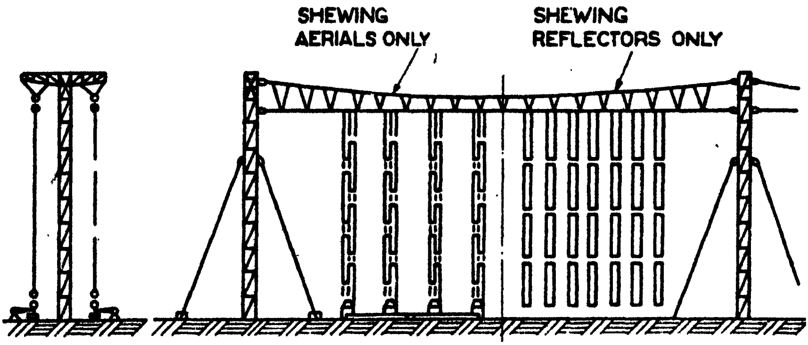
for long distance work, and the type of aerial chosen will depend largely upon the feeder system adopted. The B.B.C. claim to have proved the superiority of the horizontal aerial over the vertical for their type of work, but other organisations using both types of aerials over long periods have found no difference. Any difference that exists may be due to the way in which absorption near the aerial site affects waves of different polarisation.

We will now describe a few typical examples of broadside arrays.

Marconi Broadside Array. The desirable features of an array discussed in the general requirements paragraph can best be attained by means of an array of separate isolated

unit radiators. As it is not economic to provide too many feed points, a system must be developed with some indirectly energised wires, which can be numerous without increasing complication.

Since the economical height of an array is limited and the zenithal polar diagram very important, it is a difficult but essential matter to provide as good a zenithal diagram as possible with a reasonable height. The Franklin uniform aerial (described on page 182) gives the best efficiency in this respect since it is the only aerial in which the phase reversing device



"MARCONI" ARRAY
FIGURE 115.

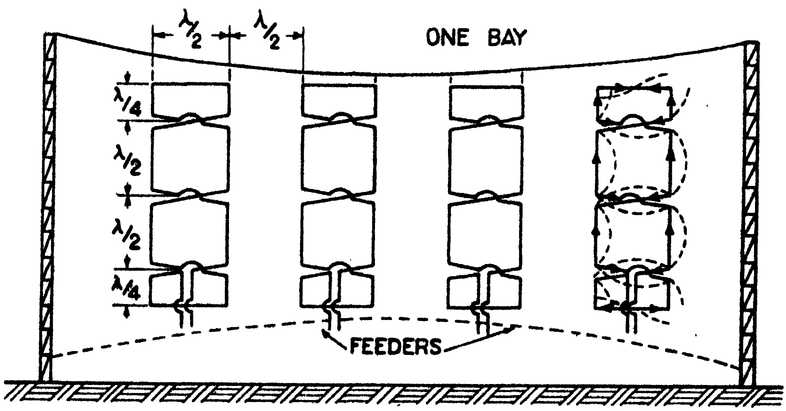
actually assists in forward radiation, and the aerial approximates very nearly to a uniform current sheet.

A sketch of a typical Marconi beam array is given in Fig. 115, aerials and reflectors being shown separated for clearness. The aerials are of the uniform type and the reflectors each comprise two, or sometimes three, parallel wires closely spaced. The effect of increasing the number of reflector units is not to improve the diagram materially, but to flatten the tuning of the array. At the same time it is found this increase of reflectors does decrease backward radiation.

The reflector aerials are usually placed approximately a quarter of a wavelength behind the energised aerials in the longer wavelength arrays, and three-quarters of a wavelength for the shorter waves, the lengths of individual aerials being adjusted to give the best diagram. There are twice as many

reflector units as energised aerials, their spacing being approximately $\frac{\lambda}{4}$ whilst that of the energised aerials is $\frac{\lambda}{2}$.

The number of aerials and, therefore, the length of the array naturally depends upon the narrowness of the beam which it is economically or otherwise desirable to produce. With one or two exceptions eight wavelengths is the maximum aperture made, and six, four and two wavelength apertures are also used. The feeder system supplying the aerials in the same phase has already been discussed in Chapter VI.



"STERBA" ARRAY.
FIGURE 116.

The Sterba Array. The Sterba Array was developed in the U.S.A., and is used on a number of the telephone links constructed by the International Telephone and Telegraph Corporation. The array is built up of a number of units of the form shown in the Fig. 116. This unit consists of a wire which, when it is supplied with current at the correct frequency, will carry a system of stationary waves. The wire is bent up in such a way that half wavelengths all of which are of the same phase, are vertical and radiate, whilst the alternate half wavelengths which would produce opposing radiations are horizontal and cancel each other out due to their proximity to each other. The current is supplied through a feeder at the centre of a half-wave portion and there-

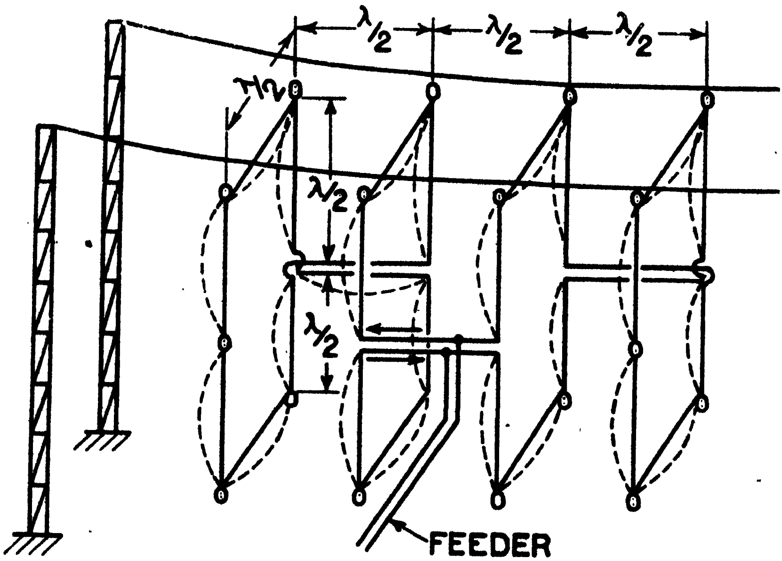
fore at a point of maximum current and minimum impedance. A number of these units are arranged side by side and supplied with current in precisely the same phase by arranging the lengths of feeder to be identical and terminating correctly.

A similar set of units is arranged approximately one quarter of a wavelength behind the first to act as reflector. Evidently, since we have here a tiered system of half-wave aerials, the radiation will be confined vertically into a beam making a small angle with the earth, and the sharpness of the horizontal beam will depend upon the number of units arranged side by side.

A useful feature of arrays such as the Sterba, which form closed metallic circuits, is that large low frequency currents can be circulated round the array to remove sleet which, if allowed to remain, might load the wires to breaking point. In countries where climatic conditions warrant it, a 50-cycle 1,000 volt supply is arranged to the Sterba array producing a current up to 200 amperes in the array wires. The heating current is supplied to the array in such a manner that the high and low frequencies can be connected simultaneously without mutual interference.

The T.W. Array. (Used by the British Post Office.)[9] Both a vertical and a horizontal type are in use, one unit of a vertical being shown in Fig. 117. Like the Sterba array, each unit consists of a long wire bent up into a suitable form so that unwanted half wavelengths cancel, due to proximity. In this case there are really two rows of aerials, in addition to the reflector (which is not shown in Fig. 117) spaced $\frac{\lambda}{2}$ apart. It will be seen that the vertical members of each row carry currents in phase with each other, but that there is a phase difference of 180° between the two rows. Hence they assist each other in both directions along a line normal to the plane of the aerials, and a reflector is still required to cut out the "backward" radiation. This reflector consists of a curtain of half-wave aerials arranged in two vertical stacks $\frac{\lambda}{4}$ behind the rear row of aerials and therefore $\frac{3\lambda}{4}$ behind the front set.

Arrays Formed of Horizontal Doublets. Several types working on this principle have been developed, notably in Germany and Holland. It was explained in the previous chapter that horizontally polarised waves are as effective for long-distance, short-wave communication as vertically polarised waves.



T·W· ARRAY

FIGURE 117.

In the particular type shown in Fig. 118 horizontal half-wave aerials are connected across the vertical parallel wire feeders at $\frac{\lambda}{2}$ intervals. The voltages at two points on the feeder will be in phase opposition, but it will be seen that the alternate half-wave aerials are cross-connected to the feeder wires so that they oscillate in phase with each other.

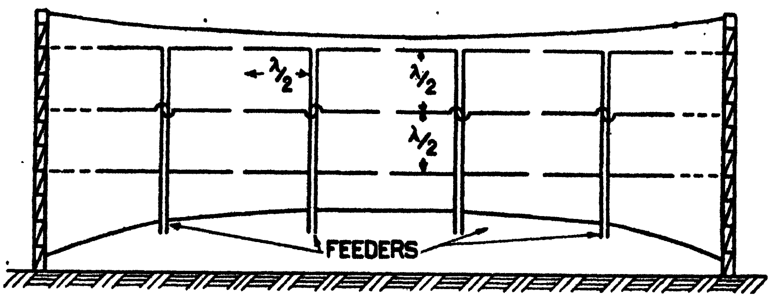
The horizontal polar diagram will be nearly the same as that given by a flat projector using vertical aerials. Regarding the vertical polar diagram, it will be seen that a distant point on the horizontal plane will be equi-distant from all the half-wave aerials and will therefore receive the maximum field

strength, whilst at a point vertically above the array the radiation from the individual aerials will cancel out.

One advantage of this design is that since the various aerials are fed through an unterminated feeder at half-wave intervals, the aerials may be regarded as fed in parallel and in consequence each aerial will share the power equally.

Swinging a Beam. It is often convenient to be able to swing a broadside array electrically to one side or the other, or alternatively to use two halves of an array to work in two different directions at the same time.

As has been explained, this can be done by changing the time-phase along the array line, and with a multiple feeder these operations are accomplished in a simple manner by lagging



HORIZONTAL ARRAY
FIGURE 118.

the current to successive aerials as desired. For instance, consider a 4 aerial beam as shown in Fig. 119; if we wish to bias the beam from normal towards the right, aerial No. 4 must be energised first and each succeeding aerial 3, 2 and 1 given a lag on the previous aerial. This can be accomplished by inserting "trombones" in the feeder tee branches as shown in Fig. 119, whose length must be correct to give the desired angle of phase lag; or the feeder can be laid out asymmetrically as shown in Fig. 120. Assume we wish to swing a beam ($\theta° = 10°$) to the right. Then if A and B (or A' and B') be two points on a feeder system Y metres apart along a line normal to the wavefront, B must be given a lag by inserting an additional length of feeder between C and B, the amount required being

$$X = Y \sin \theta$$

It is easy to understand how accurately phasing can be accomplished, as even at the highest frequencies quite an appreciable length of feeder may represent only a degree or so phase shift. For instance, if the wavelength is 20 metres, then 20 centimetres length will represent no more than 3·6°.

Broadside arrays all concentrate energy approximately in a direction normal to the plane of array line.

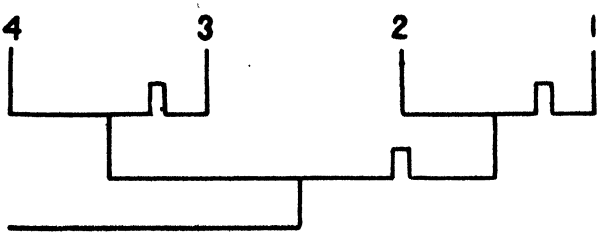


FIGURE 119.

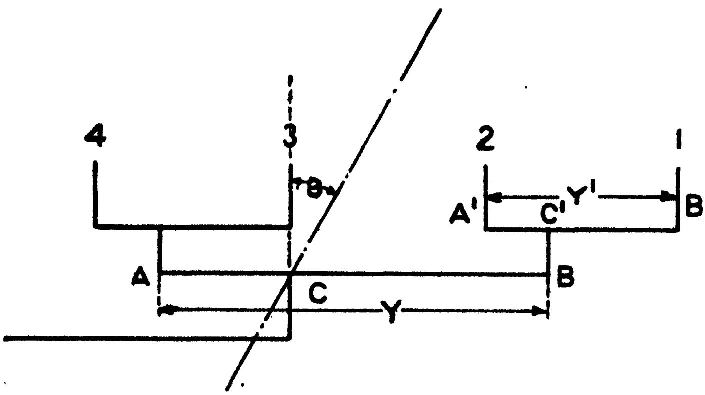


FIGURE 120.

The shape of the horizontal polar diagram depends upon the width of the array, and can be made as sharp as necessary without undue difficulty or expense. The sharpness of the zenithal diagram depends upon the height of the array, which must therefore be considerable if it is desired to concentrate the radiation into a small zenithal angle. At the same time the fact that we can control the width and height of the array independently signifies that we have a means of controlling directivity in two planes. This may be desirable since low angle rays are chiefly of use for long distance communication

and higher angle rays of more value where the communication distance is short. Further, it is easy to feed a broadside array at a number of points, and so keep a uniform current distribution and make all parts of the array equally effective as mentioned previously.

There is a limit to the effective aperture that can be used in the case of an array, on account of the wave front becoming distorted. With a broadside type of array about 8 to 10 wavelengths have been used, but the present tendency is to use arrays having apertures of not more than 5 to 6 wavelengths. All the broadside arrays described are equally effective for transmission or reception purposes, as their effective radiation efficiency is high.

" End-Fire " Arrays. We have already shown theoretically that by making the time-phase of radiators in a single line equal to the space-phase, a maximum energy concentration will be produced along a single array line without a reflector curtain, and in this case since the concentration in all planes is a function of length, the array height can be small, and the prime cost in consequence much less than the broadside array.

Although the " broadside " array gives a sharper, horizontal polar-diagram than the " end-fire " type for a given horizontal length, if the solid polar diagrams are compared, both may be equally effective in concentrating energy. In the same way that the maximum useful aperture of a broadside array is limited so the useful length of an " end-fire " array is restricted to an even greater extent due to multiple ray paths, and end-fire arrays are seldom made greater than 4 wavelengths.

Different forms of "end-fire" arrays are now in use, but in many ways the results obtained with them are disappointing compared with the broadside types. It is possible this may be because of one or two reasons :

1. Most "end-fire" arrays are terminated, whereas broadside arrays are all of the stationary-wave type. Terminated arrays are generally inefficient for transmission purposes because a percentage of the total power into the arrays must of necessity be wasted in the terminating resistance.

2. All end-fire arrays are fed from a single feed-point at one end.

As regards the second feature, all "end-fire" arrays are fed from one end, as it is such an elegant solution to the problem of making the time phase equal the space phase. The authors are inclined to believe, however, that this solution, although elegant, is ineffective and that not until designers have the courage to adopt a multiple-feed system, which will level up the distribution of energy along the array line, will the end-fire array be able to compete with its broadside counterpart.

R.C.A. Arrays Using Long Radiating Wires.[12] We have seen that a long "harmonic" aerial directs its main radiation at a decreasing angle with its own axis as the length is increased.

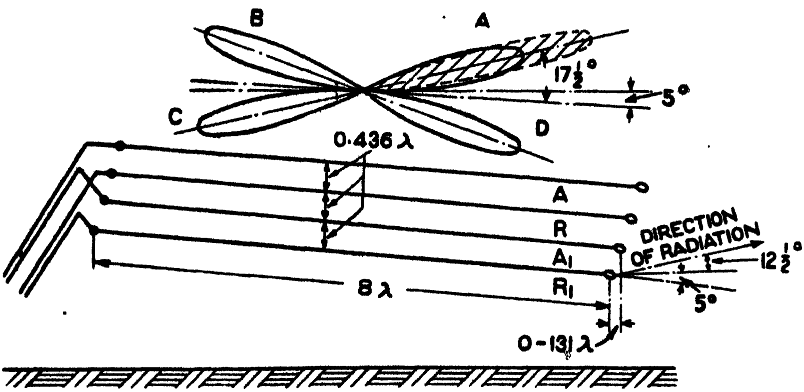


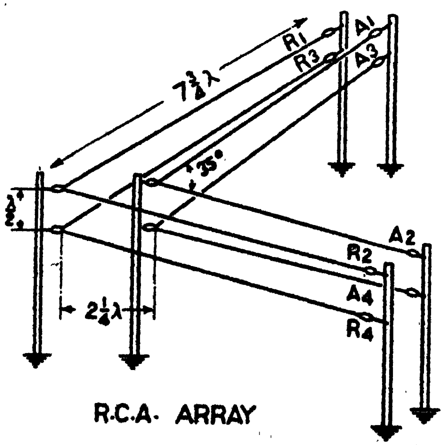
FIGURE 121.

For instance, the polar diagram of a wire eight wavelengths long is characterised by a main loop making an angle of 17½° with the axis and a number of small tails. The same diagram will be obtained for all angles perpendicular to the axis of the wire, and hence the solid polar diagram consists of two main cones of radiation concentric with the axis of the wire together with a number of subsidiary cones. This property of a long wire forms the basis of two arrays briefly discussed below.

Suppose we erect a wire A at an angle of 5° with the ground, then A, B, C, and D (Fig. 121) depict the zenithal sections of the main cones of radiation, the small tails being omitted. The maximum radiation from section A is then at an angle of 12½° with the earth's surface, this being considered the most useful zenithal angle to use for communication purposes on the particular system considered.

If a second wire A_1 is erected, parallel to the first, and fed in opposite phase, such as from a transmission line, the total field will fall away to zero in the plane normal to that containing both wires; because any distant point is equidistant from each, and the wires are fed in phase opposition. Thus the polar diagram shown is true only for the zenithal plane, and some measure of horizontal directivity has been obtained.

It is necessary still to get rid of three of the four main lobes from the diagram. It will be seen from Fig. 121 that the wires are staggered, the amount of displacement endwise being such that in a direction B, or D, any point of one wire is in phase opposition to the corresponding point on the other. The effect of this is that in these directions a zero field is obtained, and these lobes are each split into two small tails. In order to get rid of the remaining lobe C, it is necessary to erect a second pair of wires R and R_1 supplied with current in phase opposition to each other, and in quadrature with the currents in A and A_1. When the current in R leads 90° on that in A, then lobe C breaks up into small tails and hence the required uni-directional diagram is obtained.



R.C.A. ARRAY

FIGURE 122.

Another arrangement of the harmonic aerial is the "V" type shown in Fig. 122. The harmonic aerials are set at an angle and fed at the centre and thus the currents in A_1 and A_2 are in phase opposition with each other, and it follows that the cones of radiation produced by each wire at an angle of $17\frac{1}{2}$° with itself will unite if the wires are folded at an angle of 35°. The maximum radiation will, therefore, be along the line bisecting the "V," but the arrangement will be bi-directional. A similar "V" is therefore erected, an odd quarter wavelength behind the first (usually $2\frac{1}{4}\lambda$), and supplied with current in

quadrature, either leading or lagging, according to whether transmission in directions *RA* or *AR* is required.

By the use of a second "V" a half wavelength below the first, the high angle radiation is reduced. It will be seen that vertical radiation would be entirely cancelled. The first type of array produces vertically polarised waves and the second horizontally polarised

Terminated "End-Fire" Arrays. Consider a vertically-polarised, horizontally-propagated wave arriving at a vertical wire connected at its base to a receiving system whose input impedance is equal to the equivalent impedance R_0 of the

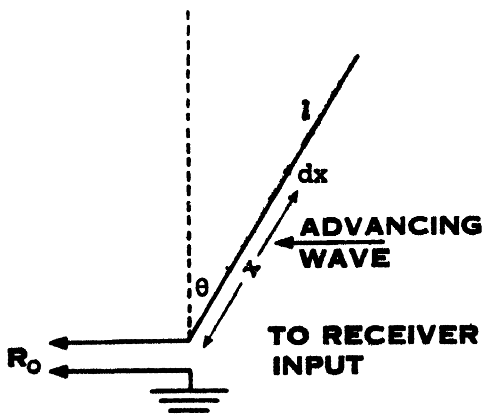


FIGURE 123.

wire, as shown dotted in Fig. 123. It will be assumed that the far end of this wire will also be terminated in a resistance equal to R_0 and therefore in our discussion we shall not consider reflection from the top end of the wire as affecting the problem. The arriving wave will induce in each elementary length of wire an E.M.F., and a current wave will therefore be initiated from each element, which will travel down the wire to the receiver input. The wire can virtually be considered as a feeder collecting energy along its length and the resultant E.M.F. at the receiver input will be dependent upon the space-phase of induced E.M.Fs. plus the time-phase due to the feeder length from each element. In the case cited the space-phase is zero and the time-phase dependent upon the

length of wire. For instance, if the wire is $\frac{\lambda}{2}$ long the space-phase is 0°, but the time-phase is such that the wave from the topmost element will lag 180° on that from the bottom element and hence the vector diagram will form a semicircle. Whereas if the wire is λ long the vector diagram will be a circle.

If now we advance the top end of such a wire into the wave by tilting the wire (as shown full line), we shall introduce a space-phase which will advance the induced E.M.F. in the topmost end of the wire to the greatest extent and the resultant vector diagram will unwrap itself. Assuming for the moment that tilting the wire does not reduce the E.M.F. induced in each element, the more we tilt the wire the more influence the space-phase has and tilting to a horizontal position would give the maximum resultant E.M.F. to the receiver, because in this position the space-phase exactly equals the time-phase and the resultant vector diagram has thereby been unwrapped from whatever shape it was with the wire vertical to a straight line ; this is true no matter what the length of wire.

Conversely, tilting the wire away from the direction of the advancing wave will wrap the vector diagram up more and more and in consequence the system will be seen to have directional properties, that is assuming there is no reflection from the far end. From previous sections it will be clear that the polar diagram will be dependent upon wire length, the longer the wire the better the diagram.

The tilting of a straight wire will not, however, be very efficient, because as it becomes more and more horizontal each element will have reduced radiation efficiency for vertically polarised waves (reception efficiency in the case of a receiver wire), the zenithal polar curve of each element having a cosine law.

R.C.A. Long Wave Array. Beverage used such a horizontal terminated wire (Fig. 124a) as a receiving aerial for long wavelengths at Riverhead, U.S.A. This aerial, which was 10 miles in length, had good directional properties but poor radiation efficiency for reasons explained above ; in fact if the received waves were exactly vertically polarised no E.M.F. would be received by such a wire, but at Riverhead the soil is very dry and sandy and this had the effect of giving the

received wave a considerable forward tilt so that a horizontal component was in evidence which induced E.M.Fs. in each element of the wire. The terminating resistance R_0 shown in Fig. 124a prevents reflection from the far end and gives the array unidirectional properties, because although waves arriving from opposite directions will build up a large E.M.F. at R_0, the energy there is completely absorbed and not reflected back to the receiver end.

R.C.A. Fishbone Array. Such an array would appear to be as suitable for short waves as for long, but since short waves are variously polarised, a modified form of array has been

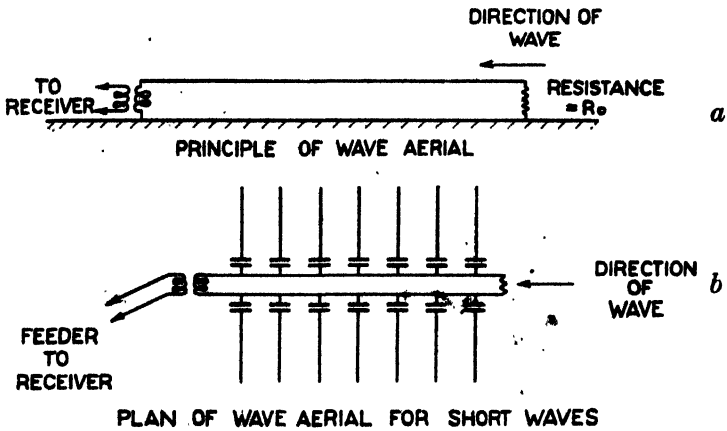


FIGURE 124.

built consisting of a two-wire transmission line of low surge impedance (about 350 ohm) to which horizontal pick-up wires are coupled as shown in Fig. 124b, and as with the original Beverage aerial, the far end is terminated by a resistance R_0 to produce a unidirectional diagram. As produced by the R.C.A. this array was designed to have flat tuning, covering a wave-range of about 4/1 efficiently, this being accomplished by using "pick-up" wires. If λ is the shortest wavelength to be received, then the spacing is $\frac{\lambda}{6}$ and the length of each wire rather more than $\frac{\lambda}{4}$. The wires are coupled to the transmission-line through capacitors of special design (Fig. 125),

and because of this small capacity in series with the wires they are electrically less than $\frac{\lambda}{4}$ and present a capacity reactance to the transmission-line at all wavelengths being received. Because the velocity of the wave along the line is lower than that of the arriving wave in space, it is not useful to make the array more than 8λ long. For a greater length, the voltages contributed to the receiver by the furthest wires would lag so far behind those from the nearest ones as to add but little to the vector resultant.

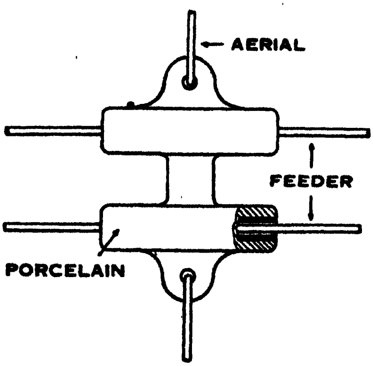


FIGURE 125.

Messrs. Cable & Wireless, Ltd., have used a modified form of this array for spot-frequency working by eliminating the coupling condensers, lengthening the pick-up wire to $\frac{\lambda}{2}$ (less 25%) and increasing the spacing between wires to $\frac{\lambda}{4}$ Since they are end-fed to the feeder line this is spaced more widely, so as to bring its characteristic impedance to 600 ohms. The array length is also reduced to about 4λ and when increased gain is desired, arrays will be paralleled.

Rhombic Arrays. Returning to a consideration of the tilted wire coupled to a receiver of input impedance R_o as shown in Fig. 123, it was mentioned that as we tilt the wire more and more into the wave, although the space phase is tending to counteract the misphase due to the wire length, the pick-up in each elemental length is getting smaller. From our previous discussion on array systems it is easy to derive an expression for the relative E.M.F. produced at the receiver for different angles of tilt, treating the wire as a feeder line to which is coupled an infinite number of elemental pick-up aerials.

Let the field strength be E volts per metre. Dividing the aerial up into a series of elemental lengths dx gives us a series of elemental generators of voltage E cos dx. These generators

are not in phase with respect to the current in the base impedance, but the phase between them is determined by summing up the phase lag given by the time taken to travel from the given element position to the base and the lead (or lag) imparted by tilting the aerial into (or away from) the source of transmission.

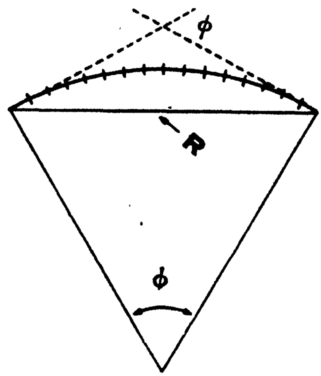


FIGURE 126.

This condition is represented by a series of vectors of length E cos dx distributed round the circumference of a circle, as shown in Fig. 126. Now the total phase angle ϕ between the first and last elemental vector is equal to the angle due to the time of travel along the wire from B to A minus the angle of lead due to the tilting of the wire into the wave. That is

$$\phi = \frac{2\pi l}{\lambda} - \frac{2\pi l}{\lambda}\sin\theta$$

The resultant voltage at A is represented by the vector R, and as has been shown previously this subtends an angle ϕ at the centre of the circle of vectors. Thus:

$$R = 2r\sin\frac{\phi}{2}$$

where r = radius of circle.

The arc $= r\phi = lE\cos\theta$

$$\text{or } r = \frac{lE\cos\theta}{\phi}$$

$$\therefore R = \frac{2lE\cos\theta}{\phi}\sin\frac{\phi}{2}$$

$$= \frac{lE\cos\theta\sin\frac{\phi}{2}}{\frac{\phi}{2}}$$

$$= \frac{lE\cos\theta\sin\left\{\frac{\pi l}{\lambda}(1-\sin\theta)\right\}}{\frac{\pi l}{\lambda}(1-\sin\theta)}$$

If this expression is differentiated to find the angle at which the resultant R is a maximum, it will be found that for each length of wire there is an angle of tilt which gives a maximum

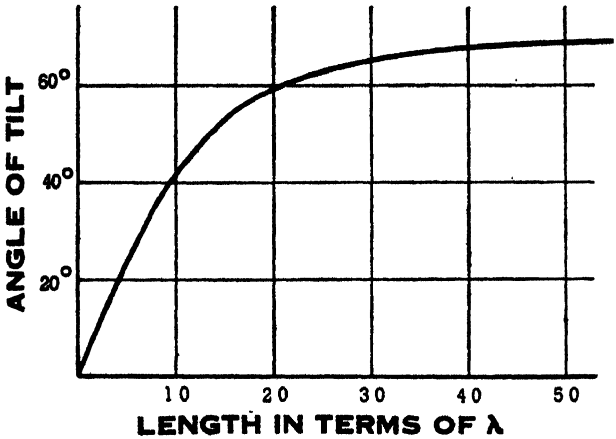


FIGURE 127.

resultant from a vertically-polarised wave and it is clear from the previous discussion that the longer the wire the greater the tilt angle necessary to obtain a maximum pick-up,

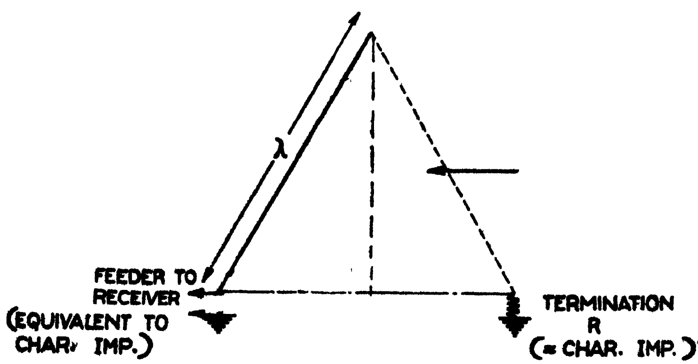


FIGURE 128.

the curve connecting tilt angle and wire length being shown in Fig. 127. Observe that for short wires the tilt angle is critical, but less so for long wires. Since the foregoing only presupposes travelling waves we must terminate the far end by a system

which will not reflect, say, by adding a second wire (shown dotted in Fig. 128), terminated in a resistance R_0.

Rhombic Array. International Telegraph and Telephone Company. This principal has been applied in the well-known Rhombic (or Diamond) aerials first produced by the I.T. and T. Company, but they are usually constructed for the reception of horizontally-polarised waves by turning the wire system to a horizontal position and adding a second pair, as shown in Fig. 129. This setting up of a horizontal rhombus at a height h above earth will, of course, modify the polar diagram and the value of received voltage from a given field

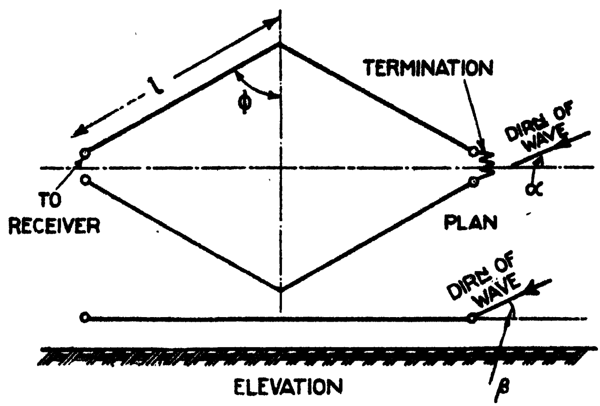


FIGURE 129.

and by small alteration of the wire lengths and angle it is possible to arrange for maximum directivity to be at any given angle β from the horizontal.

An interesting and useful feature of the Rhombic aerial, which is of particular value in reception work, is that the polar diagram varies slowly with λ and it is therefore suitable for operating over a wide wave-band. This will be evident from the consideration of the diagram showing the angle of tilt, which changes but slowly for long wires such as would be used. For a normal array, a 30% change of incoming frequency would only produce a drop of some 2 db. in the input to the receiver.

Rhombic Transmitting Aerials. The single rhombus aerial just discussed is really only suitable for reception pur-

poses because of its low radiation efficiency. If used for transmission purposes, some 40% to 50% of the power input would

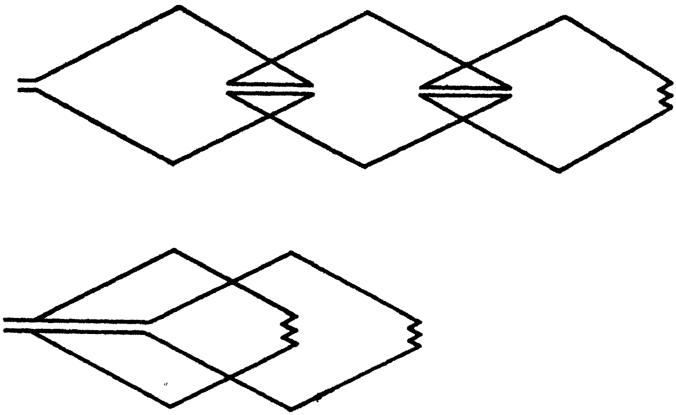


FIGURE 130.

be wasted in the terminating resistance. It is found, however, that rhombics may be grouped in series or parallel in such a way that their combined directivity is maintained (or even

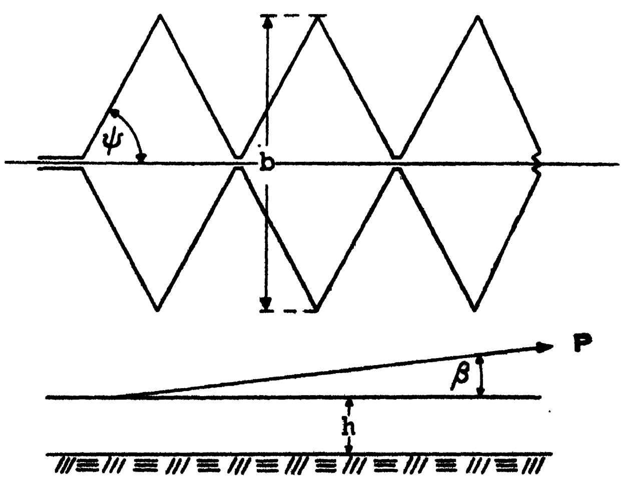


FIGURE 131.

improved) for a given area occupied by the complete array and the radiation efficiency can be brought up to more than 90%. Various forms of such grouped rhombics have been

evolved by the Marconi Company, three of which are shown in Figs. 130 and 131, the arrangement of Fig. 131 being the form most used. As with a single rhombus, the dimensions can be arranged so as to obtain a maximum directivity at any angle to the horizontal. It should be remarked, however, that such a system is now critical as to wavelength.

Marconi-Franklin Series-Phase Array. In its simplest form the series-phase array consists of a wire folded into a number of loops connected by horizontal wire lengths as shown in Fig. 132, suspended either vertically or horizontally, the dimensions of the loops and the spacing being dependent upon the type of diagram required. In general the most commonly adopted arrays are made with loops approximately one-quarter wavelength long spaced a similar amount, the

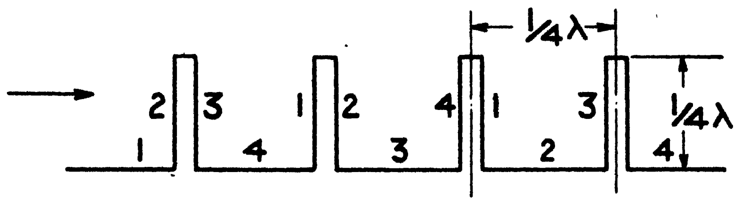


FIGURE 132.

length of the array line being dependent upon the directivity desired. An array line will be fed from one end usually through a short length of non-radiating feeder coupled to a normal concentric tube main feeder, the remote end of the array generally being terminated by a resistance equal to the characteristic resistance of the system, which approximates to 300 ohms.

As will be seen later, the loops perform two separate functions; to act as radiators; and what is as important, to determine the time phase of current between loops.

Consider an earthed vertical single wire aerial. When excited from the base, a stationary wave is formed, by a wave W_1 travelling up the wire and a similar reflected wave W_2 travelling back. We could imagine wave W_1 travelling up the left-hand edge of the wire and the same travelling wave returning down the right-hand edge of the wire, and because at all intervals of time the instantaneous values of the current

waves I_1 and I_2 at the top are equal but opposite in direction, they form a node of current here. At other points down the wire the instantaneous amplitude of I_1 and I_2 are not always equal and if their values are traced out in time they will be found to form a stationary wave with current antinode at the base when the wire is one-quarter wavelength long. However short or long this wire may be, a stationary wave will be formed by these two travelling waves with a node of current at the top end and current value at the bottom appropriate to the length of the wire. Accompanying the current stationary wave is a voltage wave in quadrature time phase with it and with an antinode at the top end.

If instead of providing a single wire we provide a loop of wire, Fig. 133, fed at the lower end, A say, this loop being part of a circuit in which a travelling wave is flowing, the wave will now travel up one wire AB and return by the second BC from which it continues on in the circuit, but provided these wires are sufficiently close together to be regarded as coincident in space from a radiation point of view, the loop may be regarded exactly as a single wire carrying a stationary wave with current node at B.

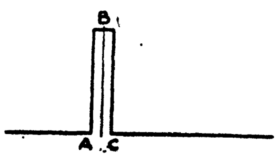


FIGURE 133.

These two travelling waves not only form a stationary wave of current with node at the top end and (if the loop is $\frac{\lambda}{4}$ or less) an antinode at the bottom end, but in quadrature time-phase with the effective current stationary wave there will be a voltage stationary wave, having an antinode at the top end and a node at the bottom end. The voltage does not reverse in sense at the top and in consequence no node is produced, whilst at the bottom of the loop the voltages are always equal but opposite in phase.

The radiation resistance of the loop will be four times the radiation resistance of a single wire for the same base current measurement in each case. This is so because a meter placed at the base of one limb of the loop is measuring current in one limb only, and this is half the effective stationary wave current at the base, as the currents add at this point. This means virtually that the effective height, and in consequence the

radiation efficiency, of this portion of such a system is high. For this reason an array built with loop radiators is equally suitable both for transmission and for reception purposes.

The radiation efficiency of the loop will not vary greatly for lengths below $\frac{\lambda}{4}$, and as will be seen shortly, the loop dimension is determined largely by how it has to act as a phasing device.

Consider Fig. 134a, which shows two radiators 1 and 2 spaced one quarter wavelength apart and connected by a feeder line. If this system is fed from a point *A*, halfway between the aerials, zero time phase is supplied to both aerials, but if we move the feed point to *B*, this automatically creates a time phase difference between 1 and 2 equal to the space phase

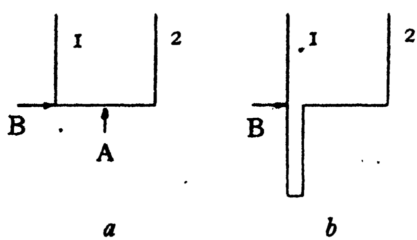


FIGURE 134.

between them, assuming the radiated wave travels to the right at the same velocity as the wave along the feeder. In this case maximum directivity is away from the feed point *B*.

Still keeping the feed input at *B* we can reverse the diagram by looping the feeder to give aerial 2 a lagging current of 90°. To do this the feeder length can be increased as shown in Fig. 134b, such that it equals (360°—90°) or $\frac{3}{4}\lambda$. If we design the loop to have $\frac{1}{4}\lambda$ sides as shown, this loop together with the straight portion of $\frac{1}{4}\lambda$ makes up the $\frac{3}{4}\lambda$, and as we have seen, if the sides of the loop are coincident in space, the loop itself will act as a radiator; in consequence we can use it not only as a phasing feeder to aerial No. 2 but to replace aerial 1. In a similar way the whole line of radiators can be replaced by loops, whose lengths are made correct to produce the required phasing between the radiating elements. This is the usual series-phase array design which therefore has maximum

directivity from its fed end, and it is clear that with this particular spacing we could not *reduce* the dimensions of the loops sufficiently to reverse the diagram, i.e. by producing a time phase equal to the space phase as the loops would then have zero dimensions.

But we can obtain this reversal by *increasing* the loop still more, namely to $\frac{\lambda}{2}$, as in this case the total feed length is then $1\frac{1}{4}\lambda$, and this gives the required time phase.

Considering $\frac{1}{4}\lambda$ loops with $\frac{1}{4}\lambda$ spacing, theoretical polar diagrams for 2, 4, 8, and 16 loops are shown in Fig. 136, these diagrams being sections of three dimensional figures having similar dimensions in all planes. An array placed close to a perfectly conducting earth would have a diagram represented by half these figures, i.e. the axis of the polar curves represents the earth plane, but on short wavelengths, due to the earth acting more nearly as an insulator, the diagrams will be modified because the radiation at very low angles will be reduced by an amount appropriate to the particular earth conditions considered.

So far the argument has assumed that the horizontal wires act merely as connecting wires and contribute nothing to radiation, that a uniform current distribution throughout the system is obtained, and that the velocity of the wave along the aerial is the same as that of the radiated wave in space.

Dealing first with the horizontal wires, these do not materially affect the radiation from the system because any radiation will be along the direction of the wire, and of negligible amount.

Discussing now the second assumption, the actual current distribution depends on the radiation efficiency of the loops, and with an accurately tuned array there is a considerable fall of current along the array. By choosing appropriate lengths the fall per loop can be minimised, but it sets a practical limit to the number of useful loops. Diagrams based on a uniform current distribution are reasonably true, however, as can be seen by considering the vector diagram of an array with tapering current. Thus Fig. 135 shows the vector diagrams for an array along which the attenuation (which will follow a logarithmic law) is such that the current in any loop is 0·875 of that in the previous one. The total vector length has been

made the same as the previous Fig. 111, and it is clear that whereas there is practically no difference between the vector diagrams for small angles near the maximum, the vector diagram with attenuation, instead of wrapping up completely, takes a spiral convolution. This means that although there is small difference in the shape of the main figure at no time is a complete minimum produced. The smaller tails, however, are not so much in evidence and since the back and side amplitudes are in any case so small compared with the main loop, the effect of the attenuated current along the array is invariably small, the dotted curve in Fig. 137 showing the polar diagram for a 16 loop array with tapering current as above. This can

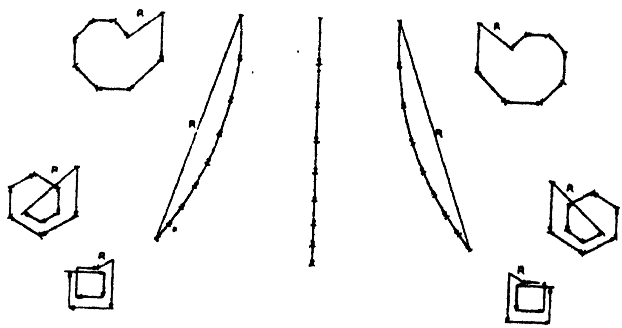


FIGURE 135.

be compared with the full curve for the same 16 loops with uniform current in Fig. 136.

A further small point is that the velocity of current along the feeder is different from that of the wave in free space, but can be allowed for by reducing the dimensions of the array. Actually array dimensions are reduced by quite a fair percentage, but for other reasons; first, to sharpen up the main diagram and secondly to control the attenuation along the array a reduction of 5 per cent. in dimensions is made for arrays above 30 metres, 10 to 15 per cent. for arrays below 30 metres and on ultra-short waves the reduction may be even more than this.

Considering Fig. 136, which shows a diagram for sixteen $\lambda/4$, loops with attenuated current, if we reduce the dimensions of the loops by 10 per cent. we obtain a diagram as shown in

Fig. 137, whereas if we increase the dimensions 10 per cent. we obtain the two wing type of diagram shown in Fig. 138. It is easy to see why the diagram alters in this way if we refer back to Fig. 112c. This shows that the "end-fire" polar curves can be imagined to be made by the overlapping of the two main loops of a bi-symmetrical diagram coming together. When the time misphase is not sufficient to add to the space phase,

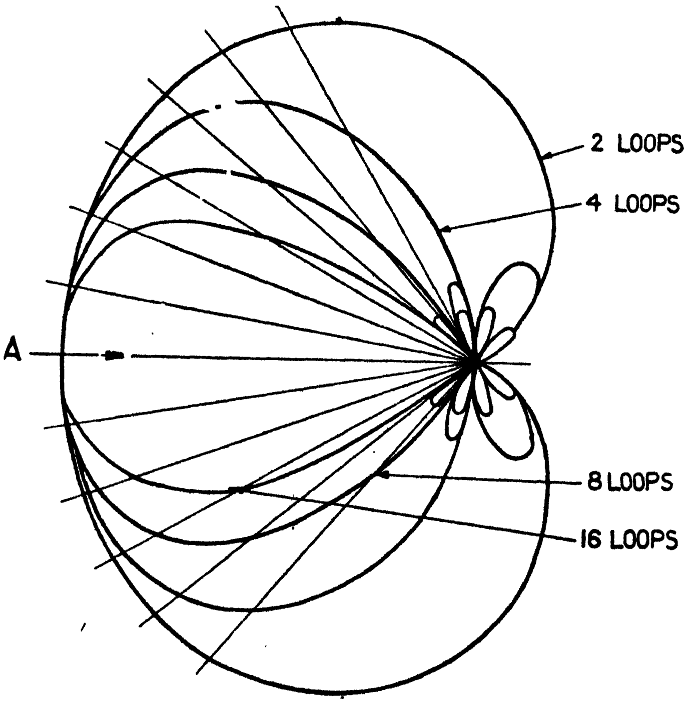


FIGURE 136.

i.e. on wavelengths shorter than the spacing, the two loops have not yet reached the overlap condition and the wing type figure is produced. Whereas when the time misphase is made greater than the space phase they have more than overlapped and their steep sides are now coming together and so produce a sharp pointed polar curve, but at the expense of greater back radiation, and we cannot carry this mistuning too far. As mentioned above, the other advantage in misphasing is that because the array is detuned, the current attentuation per loop

is less and we thereby can control the current distribution. Instead of detuning each loop the same amount we can control the current value best by varying the dimensions of each loop, detuning the first most and each succeeding loop less and less.

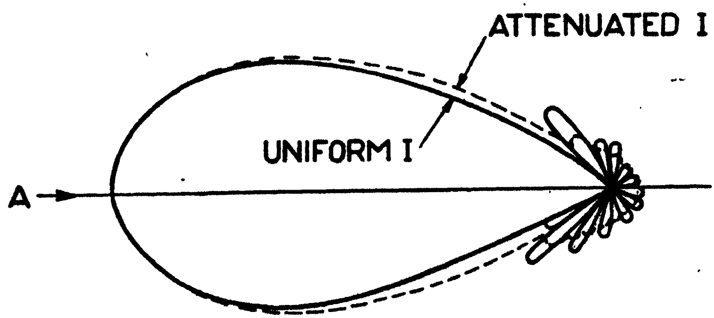


FIGURE 137.

One interesting point in connection with "end-fire" arrays generally is that practical experience with them has led to the suggestion that the transmitting and receiving systems are not strictly reversible. This has resulted in the building of larger arrays for transmission than for reception. That is to

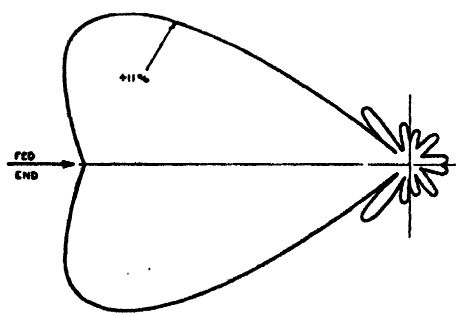


FIGURE 138.

say whereas a 15 element transmitting array is used, similar arrays acting at a receiving station have been found to be less effective and for this reason receiving arrays of the series phase type are usually kept down in length to 10 or 12 loops.

Energy Gain of Arrays. If the solid polar diagram of a directional array is known and that of a single aerial is also

known, then by integrating these polar diagrams we can find the power which each would use in order to produce the same field strength in the required direction, and this power ratio (expressed usually in decibels) will be the gain of the array over the single aerial.

Franklin[1] estimated originally that the energy gain of a broadside array system would be 9·6 per square wavelength of aperture surface compared with a half-wave aerial. The easiest comparison to make is to take one plane at a time and compare the gain of the array with one of the aerials which go to make it up,

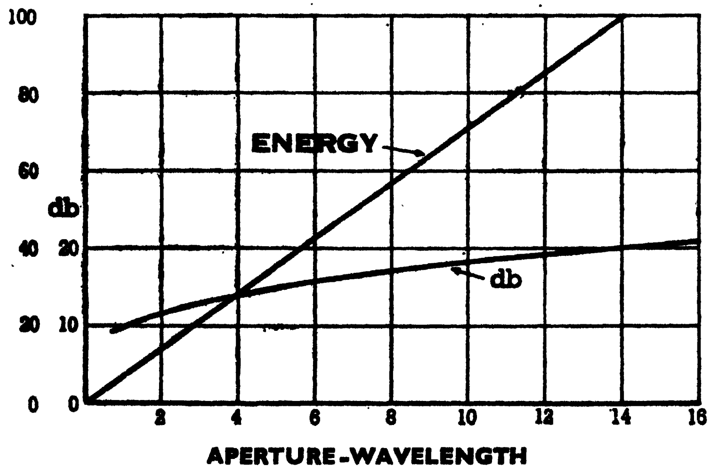


FIGURE 139.

as we may suppose in this case that the vertical polar diagram is the same for both, since the same aerial height and quality of earth is involved.

T. L. Eckersley, Green, and Southworth, treating specific cases, have produced results which are in fair agreement with Franklin's, and the curve of Fig. 139 gives average values for the gain of an array of different apertures.

It will be seen from this diagram that the energy gain is directly proportional to the array aperture. Thus a 6λ array has an energy gain of 43 (16 db.) and a 12λ array a gain of 86 (19 db.). Accordingly, if we add a 6λ array to an existing array of 6λ, we shall double the energy gain. If, however, we have an array at both transmitting and receiving ends, the

total gain of the system is now the product of the array gains, not the sum. This can best be seen by referring to the level diagram in Fig. 140. Consider an output power of 1 kW, i.e. 60 db. above a datum level of 1 mW, and an ionosphere attenua-

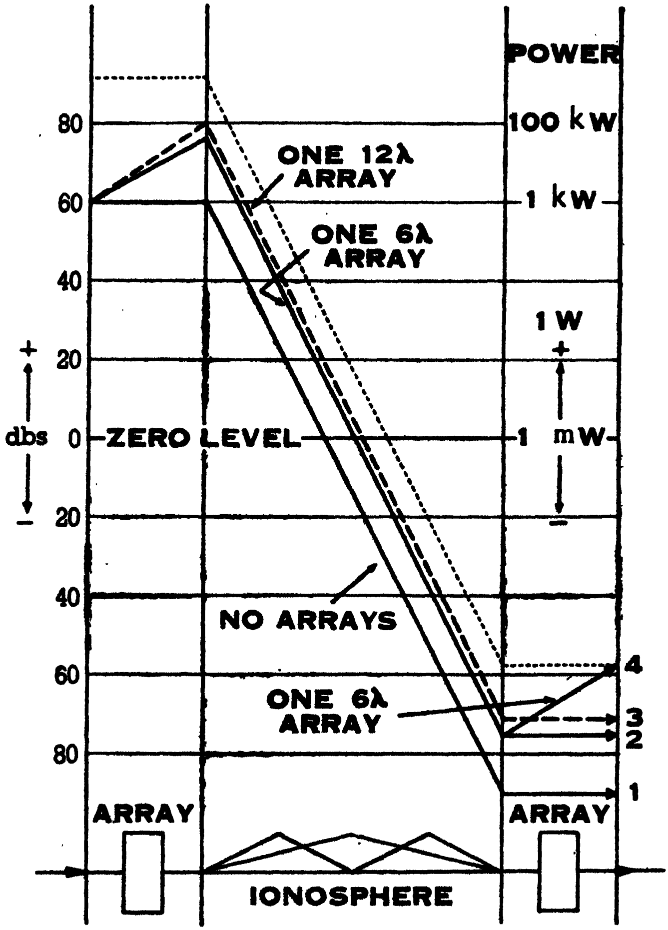


FIGURE 140.

tion of 150 db. With no arrays, the level of received signal will be -90 db., shown by curve 1. If we use at the transmitter an array having a 16 db. gain, we raise the level at the transmitting end, and therefore at the receiving end, by 16 db., as shown by curve 2. Had we increased the dimensions of the

transmitting array to 12λ, this would have increased the level at the transmitting and receiving end only by another 3 db., as shown by curve 3. On the other hand, had we left the dimensions of the transmitting array at 6λ, raising the level by 16 db., and used a 6λ array at the receiving end, the output level to the receiver would be raised by another 16 db., i.e. the total gain would have been 32 db. If omni-directional aerials had been used at both ends, it would have been necessary to raise the transmitter power to 1200 kW, to obtain the same level of received signal, as is shown by tracing back the dotted line on Fig. 140.

It might be expected that calculated values for the gain of an array system would differ considerably from the experimental figures since a number of uncertain factors are involved, but surprisingly enough, calculated and measured gains are in very fair agreement.

It is found, however, that the effective gain of an array does not remain constant but depends upon ionosphere conditions. This is because the full gain can only be obtained when we deal with a perfectly-uniform wave front, as it is only then that the phase relationhips in the various aerials are correct. Measurements taken over long periods indicate that only for a small percentage of the time is the full gain of the array obtained, Fig. 141 showing the average of some results made by Bruce on a short wave transmitter in Great Britain as received in America.

The Steerable Antenna (M.U.S.A.). It has been seen that most arrays are designed to provide a fairly-sharp zenithal polar-diagram having a maximum at a small angle from the horizontal in order to receive mainly the low-angle rays which have made the fewest "hops" and which are usually the strongest. The disadvantages attendant upon reception of too many of the rays have already been discussed (see page 114). It is evident that if the zenithal polar-diagram is to be fixed it cannot be too sharp, since the rays which are most prominent and useful vary in their angle of arrival. The M.U.S.A. system, which has been developed by the Bell System Laboratories and is shortly to be in permanent use at both ends of the transatlantic telephone service, employs a long line of horizontal rhombic arrays and is capable of very great directivity in

both planes. The vertical polar diagram is, however, instantly varied by electrical adjustments at the receiver and may therefore be adjusted to suit the varying conditions.

An outline diagram of the arrangement is shown in Fig. 142. Let us concentrate our attention upon a ray arriving

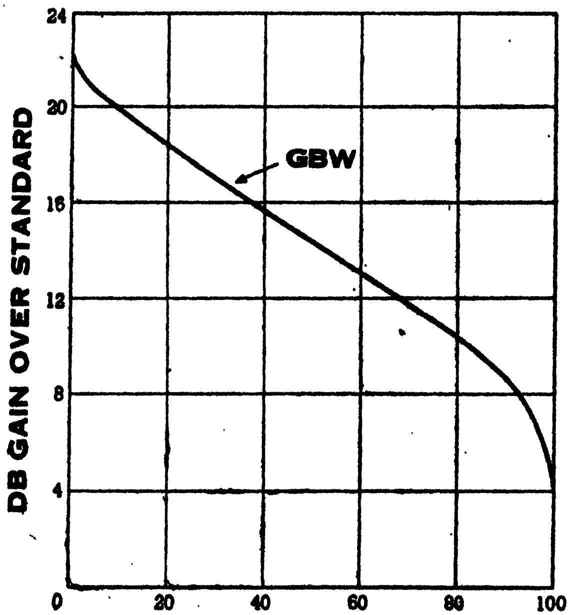


FIGURE 141.

at some angle θ. This will arrive at (2) at a time $\dfrac{d \cos \theta}{c}$ secs. later than at (1) (where c is the velocity of the wave in space). The distance along the feeder from (2) is d metres less than from (1) however, and hence, if the velocity of the wave along the feeder is v, the voltage at the receiver provided by (2) will lead on that from (1) by $\dfrac{d}{v} - \dfrac{d \cos \theta}{c}$ secs. and the phase angle between the voltages is therefore

$$2 \pi f \left(\frac{d}{v} - \frac{d \cos \theta}{c} \right) \text{ radians.}$$

Similarly, the voltage from the nth aerial will lead on that from (1) by

$$2\pi f\left(\frac{d}{v} - \frac{d\cos\theta}{c}\right)(n-1) \text{ radians.}$$

If we now provide phase-shifting arrangements in each feeder at the receiver end, we can compensate for the phase differences so that all the voltages add up in phase and produce the maximum possible resultant. This compensation will, however, only apply to the ray at the angle θ and hence this ray has been selected from others which may be present,

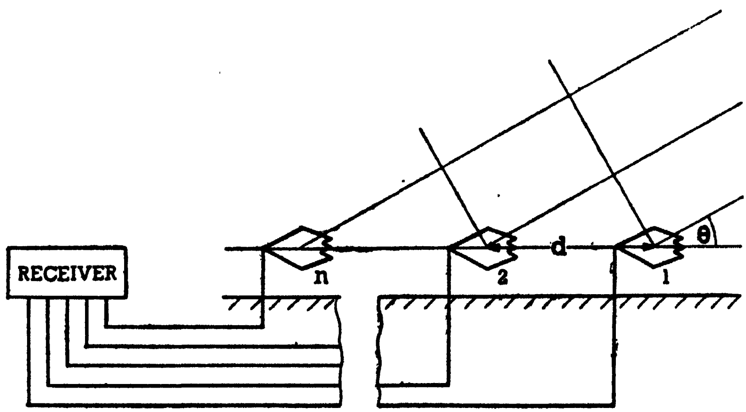


FIGURE 142.

and evidently, by altering the phase-shifts, any ray can be selected. Two or more separate phase-shifting arrangements can be provided so that several rays can be selected and later combined, though further phase adjustment will be necessary before combination because the different rays will have travelled different distances.

It will be seen that the amount of phase-shift required depends upon the frequency of the signal, in the simple system we have described. For this reason and also because it would be very difficult to control the phase-shift at short wave frequencies, in the actual equipment used the phase shifting is carried out in a later stage of the receiver where the frequency is fixed. The method used is described in outline on page 494.

Arrays for Ultra-Short Waves. The types of array which have been discussed are used for waves from, say, 3 to 10 metres as well as for short waves and it naturally becomes easier and cheaper to produce a very sharp polar diagram as the wave length is reduced.

For waves of the 1 metre order there would be great difficulty in carrying out the adjustments required in order to obtain

FIGURE 143.

correct phasing and termination. Losses in, and radiation from, any complicated feeder system are also likely to be serious. It is, therefore, much simpler to use some form of parabolic reflector, with a single aerial at the focus. Marconi and Mathieu employed the arrangement shown in Fig. 143 on a 50 cm. wavelength. This gives a very sharp zenithal polar-diagram, but the horizontal polar-diagram is merely that of a $\frac{\lambda}{2}$ aerial and is, therefore, broad. To sharpen up the

horizontal diagram, if necessary, other parabolic arrays are set up alongside, each being fed from its own transmitter. This design avoids difficulties due to windage which arise if a solid structure of requisite dimensions is used. For the still shorter 17 cm. circuit across the Straits of Dover, Standard Telephones and Cables, Ltd., employ arrays which closely follow optical principles. An aluminium, paraboloid reflector of 18λ aperture is used, with a $\frac{\lambda}{2}$ aerial at the focus, as shown in Fig. 144. A hemi-spherical mirror is mounted in front of the aerial, so that the energy leaving the aerial in a forward direction shall be directed back on to the main reflector and hence into the beam. In order to ensure that the radiation redirected in this way adds in phase with the remainder, the diameter of the hemisphere must be a multiple of $\frac{\lambda}{2}$.

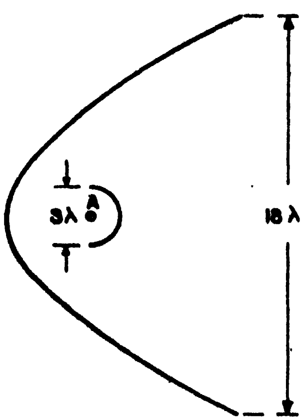


FIGURE 144.

The gain of the whole arrangement is about 30 db., and the sharpness of the beam such that turning the reflector through 3° brought the received signal down by about 10 db.

Yagi and Uda have evolved a different type of directive system, which they have named a "wave canal" (Fig. 145). An arrangement of three (or sometimes, five) parasitic aerials form what is termed a "trigonal reflector," which strengthens the radiation along the line of "directors." It was seen (p. 207) that by adjustment of tuning and spacing a parasitic aerial could be made to act as a reflector or director —Yagi used directors having a natural frequency higher than that of the wave being transmitted and found that a line of such directors, suitably tuned and spaced, produced a sharp polar-diagram with its maximum along the director line.

Such an array could, of course, be used on any short wavelength but was actually employed on 5 or 6 metres.

With wavelengths below one metre, an alternative to the parabolic reflector is the directive horn, which is similar in principle to the acoustic horn.

Such horns are particularly suitable for use as the radiating element at the end of a dielectric wave-guide and the figure

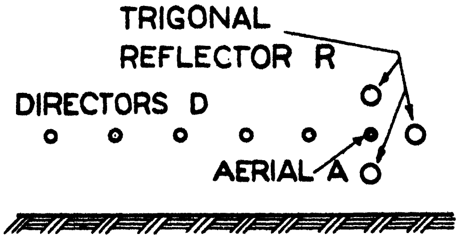


FIGURE 145.

shows an arrangement (due to Southworth) suitable for a 23 cm. wavelength, when the H_1 type of wave is propagated along the cylinder. The polar diagram is seen to compare favourably with the polar curve of a good array system.

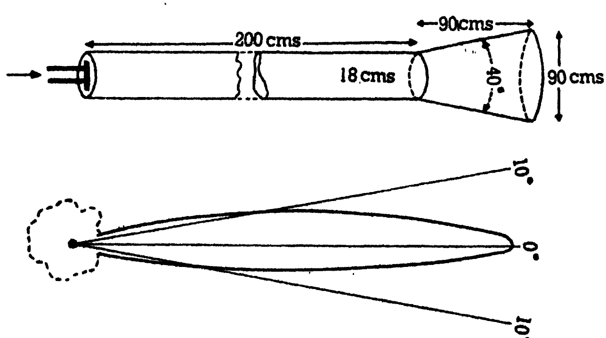


FIGURE 146.

SELECTED REFERENCES

1. *The Beam System.* G. Marconi. *Proc. Roy. Soc.*; July, 1924.
2. *Radio Communications.* G. Marconi. *Proc. Roy. Soc.*; December, 1924.
3. *Calculation of Polar Curves of Extended Aerial Systems.* E. Green. *W.E.*, Vol. 4; October, 1927.
4. *Theoretical Investigation of Phase Relationships in Beam Systems.* Wilmotte and McPetrie. *J.I.E.E.*, Vol. 66; September, 1928.

5. *Radiation Resistance of Beam Antenna.* Pistolkors. *P.I.R.E.*, Vol. 17 ; March, 1929.

6. *Action of Reflecting Antenna.* Palmer and Honeyball *J.I.E.E.*, Vol. 67 ; August, 1929.

7. *Short Wave Directive Antennae.* Ullrich and Fairbank. *Elec. Com.* ; February, 1930.

8. *Factors Affecting Gain of Directive Antennae.* Southworth. *P.I.R.E.*, Vol. 18 ; September, 1930.

9. *Beam Arrays and Transmission Lines.* Walmsley. *J.I.E.E.*, Vol. 69 ; February, 1931.

10. *Directional Transmitting Systems.* Sterba. *P.I.R.E.*, Vol. 19 ; July, 1931.

11. *Developments in Short-Wave Directive Antennae.* Bruce. *P.I.R.E.*, Vol. 19 ; August, 1931.

12. *Development of Aerial Arrays.* Carter, Hansell and Lindenblad. *P.I.R.E.*, Vol. 19 ; October, 1931.

13. *Diversity Receiving Systems of R.C.A. Communications, Inc. for Radiotelegraphy.* Beverage & Petersen. *P.I.R.E.*, Vol. 19 ; April, 1931.

14. *Transoceanic Telephone Service.* A. A. Oswald. *Bell System Technical Journal*, Vol. IX, p. 270.

15. *Horizontal Rhombic Antennas.* E. Bruce, A. C. Beck and L. R. Lowry. *P.I.R.E.*, Vol. 23 ; January, 1935.

16. *An Unattended Ultra-Short-Wave Radiotelephone System.* N. F. Schlaack and F. A. Polkinghorn. *P.I.R.E.*, Vol. 26 ; November, 1935.

17. *The Wave Antenna.* H. H. Beverage, C. W. Rice and E. W. Kellog. *Trans. A.I.E.E.*, Vol. 42, p. 215 ; 1923.

18. *A Multiple-Unit, Steerable Antenna for Short-Wave Reception.* H. T. Friis and C. B. Feldman. *P.I.R.E.*, Vol. 28 ; July, 1937.

19. *Circuit Relations and Radiating Systems.* Carter. *P.I.R.E.*, Vol. 20 ; June, 1932.

20. *Directional Antennas.* Brown. *P.I.R.E.*, Vol. 25 ; January, 1937.

21. *Developments in the Beam Array.* Ladner. *B.A. Meeting.* 1936.

22. *The Series-Phase Array.* Ladner. *Marconi Review.* September–December, 1937.
23. *The "Sectoral" E. M. Horn.* Barrow and Lewis. *P.I.R.E.*, Vol. 27; January, 1939.
24. *Theory of E.M. Horn.* Barrow and Chu. *P.I.R.E.*, Vol. 27; January, 1939.
25. *Metal Horns as Directive Receivers.* Southworth and King. *P.I.R.E.*, Vol. 27; February, 1939.

CHAPTER IX

PUSH-PULL

The name "push-pull" is given to those circuits impulsed by two valves arranged differentially. Of course it is not necessary to employ two valves to obtain a push-pull action, nor does the name adequately describe the circuit to which it is applied, for it is possible, with small modifications of such a two-valve circuit, to produce a variety of effects, some of which are not push-pull in character. Push-pull circuits have a number of uses on short waves, both in transmitting and receiving work, and an analysis of one stage of a general form of amplifier will indicate their operation for most cases, the special circuits being treated separately in other chapters.

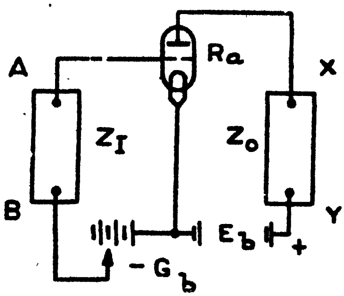


FIGURE 147.

Let us consider first a single-valve, amplifying stage as shown in Fig. 147, where AB is an input circuit of suitable design; R_a the valve, and XY a suitable output circuit. Suppose an E.M.F., E_{g1} to be impressed across AB. If the static grid-bias is such that the valve can operate wholly on the straight part of its characteristic, which implies freedom from grid current and bottom bend curvature regions (as shown in Fig. 148), the resulting anode current will impulse XY to give distortionless output, whereas if the bias is set sufficiently negative the impulse will be in one direction only, for in this case the feed current will be partially or wholly rectified, as shown in Fig. 149.

The first type of amplification known as Class A will give a push-pull action on the output, and would reproduce faithfully

the waveshape of the input signal, whereas the second type, known as Class B, is non-linear, and this would naturally lead to the introduction of harmonics in the output. Only as long as the characteristic of the system is wholly linear will the effect be purely push-pull; thus with a single valve (or paralleled valves connected as for a single valve) adjusted for push-pull working a quiescent value of anode feed must be provided of value at least equal to the peak A.C. required, in fact more, because of the bottom bend of the characteristics. Besides being wasteful, this D.C. feed may not be desirable if it is large and has to flow through the windings of the output circuit.

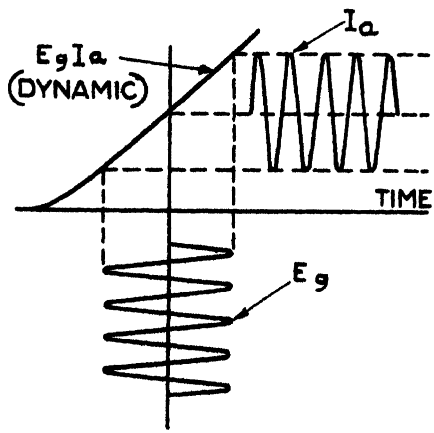


FIGURE 148.

We will now consider the addition of a second (similar) valve to the above circuit, the arrangement being as shown in Fig. 150, which is the general form of push-pull circuit. If we have similar input and output circuits to those used with the single valve, then in order to connect in the second valve it will be necessary to bring out centre tappings from each circuit as shown, the input being connected to grid-bias and filament, and the output being connected to H.T. and back to filament. This leads to an arrangement being produced which is symmetrical in every way, including electrostatic capacity effects from all parts of the circuit to earth.

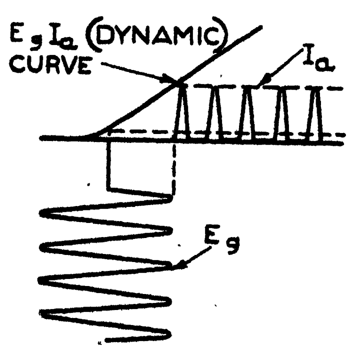


FIGURE 149.

Let us examine the operation of this push-pull circuit, assuming that the valves are set on the straight part of the characteristics so that the feed through each is equal, " a," say, as shown in Fig. 151a. Since the source of H.T. is connected to Z, the centre point of the output circuit, the main feed (of amplitude " 2a ") divides at this point, and the feeds to each valve flow in opposite directions through the output circuit. Thus as long as these feeds remain equal or change equally no A.C. output can result across XY. But any inequality of feed to the valves results in output, and a fall of current through one valve directly assists a rise of current through the other, since the feed currents are flowing in opposite directions through the output.

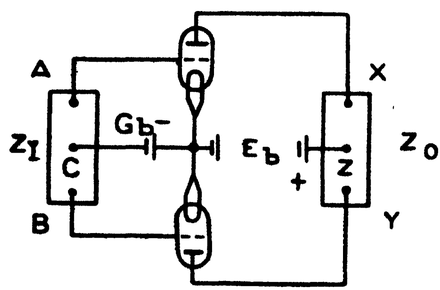


FIGURE 150.

Turning to the input side and considering the application of a signal, it will be seen that the voltage of each valve can be only half the total input voltage, and hence, from a given available E.M.F., each valve of a push-pull combination can get but half the input volts a single valve could get. Any input to AB acts differentially on the grids of the two valves, however, the voltage on grid 1 rising less negative as that on grid 2 rises more negative, and vice versa. This means that the feed to valve 1 will flow in opposite phase to the feed to valve 2, and, as explained above, such effect will be additive in the output circuit. If the valves are set on the straight part of their characteristics, each valve will have a push-pull action on half the output. Thus in Fig. 151a, if $E_g I_a$ are the dynamic curves of valves 1 and 2, the signal voltage applied to input AB is as shown by E_{g1}, E_{g2}, and this voltage causes anode currents to flow in antiphase, as shown by I_{a1}, I_{a2} (full line). But since I_{a1}, I_{a2} act differentially on XY, in order to get the resulting output we must reverse one of these currents, as shown by I_{a2} (dotted line), and the resulting current is as shown by I_o.

If the valves are each biassed negative so as to eliminate

the D.C. feed, we still get a push-pull action, provided the static bias is above the bottom bend; for in this case No. 1 valve will "push" for the first half cycle of signal, when its grid goes less negative, and No. 2 valve will "pull" for the

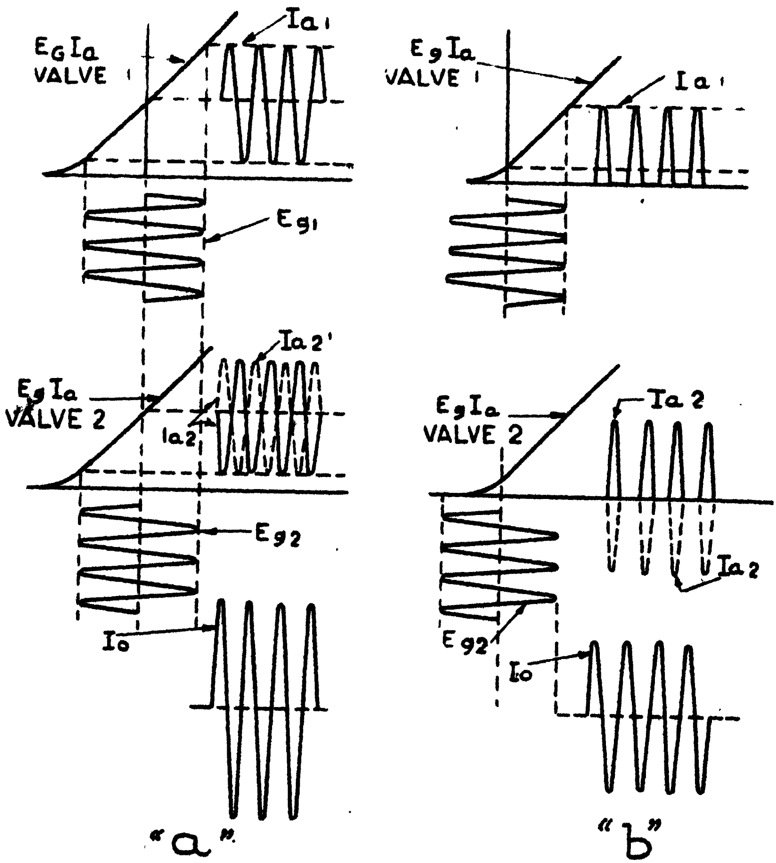


FIGURE 151.

second half cycle of signal, when its grid goes less negative. The fact that there is partial or complete rectification of the feed to each valve will make no difference to the push-pull action (provided the total effective characteristic remains substantially linear), except the amplitude of output will be reduced, as shown in Fig. 151b.

It is, of course, essential that each valve shall have the same constants, and, further, we cannot bias the valves to the cut-off point, but only above the bottom bend, so that the total characteristic is linear. If the swing on each grid is limited to a value which avoids the grid current region such an amplifier is often called Class AB, because its overall performance comes under Class A, but each valve is operating under Class B conditions.

If the valves are set back beyond the cut-off point, the feed no longer follows the signal waveshape, and harmonics are introduced into the output; with this difference from the single valve circuit, namely, that no even harmonics are present, since the impulse from the second valve is always in such a phase as to prevent them.

Before continuing it will be desirable to make a comparison of the amplifying properties of a straight single-valve circuit with the push-pull arrangement, considering two cases, firstly, with a definite input signal, and secondly where one has unlimited input.

Consider an amplifier having a total gain of 30 db. This is relative gain, the actual output depending upon the voltage of input, which in this case is the whole of the incoming signal. If we assume this to be 1 millivolt, our total actual output voltage is 31·6 mV.

Considering the push-pull arrangement, if we assume our output circuit is changed to produce an impedance across each valve similar to that used with the single valve, each half of the circuit can give as much stage gain as the single valve. Thus one might imagine that if each half can give the same gain (say 30db), the total gain will be doubled, since they are both operating on the output. But the actual output still depends upon the actual voltage applied to each grid, and since this can only be half the applied signal voltage, the actual output is but the same as with the single-valve arrangement.

From a given available input, then, the push-pull circuit gives no greater amplification than a single-valve system.

If, however, we can provide unlimited input, such as in a transmitting circuit where large power output is the first consideration, the case is different. Here we are primarily interested in power-conversion efficiency from D.C./A.C., and

being able to obtain this with a minimum emission current. If we have a single valve whose available emission, excluding bending of characteristic, is "$2a$" say, then to obtain undistorted output of a signal we should need to set the valve to pass a steady feed of "a," and the valve limits the output current to a peak amplitude of "a." To obtain this change of current and voltage a grid swing of E_g will be necessary, and with such an arrangement the maximum theoretical conversion efficient cannot exceed 50%.

Turning to the push-pull arrangement, if each valve of the push-pull circuit is biassed to near rectification, each valve has an available emission of "$2a$," hence the total available emission is "$4a$," and we can obtain undistorted output to a peak amplitude of "$2a$," provided the input grid swing is increased sufficiently, and each valve can now operate to a theoretical conversion efficiency of 72%.*

Exactly the same output could have been obtained, of course, had the two valves been paralleled. For the total available emission is still "$4a$," just as the push-pull case, only now we would need to pass a steady feed of value "$2a$" through the valves in the quiescent condition, but the conversion efficiency of each valve is still only 50% theoretical, and therefore valves having a greater anode dissipation would be required. On the other hand, the same output could have been obtained with approximately half the push-pull input voltage, because paralleling the valves improves the effective mutual conductance, whereas putting them in push-pull does not.

Rectification. It is fairly obvious from what has been said that so long as a push-pull system remains symmetrical no rectification of waveform can take place. Any waveform applied to the input AB results in amplification at the output XY, and the static biassing of the valve grids will not alter this push-pull action, although it may introduce distortion. We may say, therefore, that with a pure push-pull circuit it is impossible to rectify, and if we have a push-pull amplifier receiving a high frequency wave whose envelope it is desired to extract, since rectification is essential for selecting the modulating component, we must consider methods available, all of them involving the unbalancing of the system.

* These efficiencies are discussed in Chapter X.

Unbalancing the Circuit. The circuit can be unbalanced by setting one valve to saturation and one to zero, but this is not usual. A better method is to connect both anodes together and to one end of the output circuit (as shown in Fig. 152) and taking the common feed to the other end. Then if both valves are biassed to the bottom end of the characteristic each valve rectifies alternate half cycles of the input and a rectified wave is produced. In this case the two half cycles follow one another and rectification is of the full wave type, as shown in Fig. 153.

It may be noted that the rectification efficiency of either method is the same theoretically, but the second arrangement

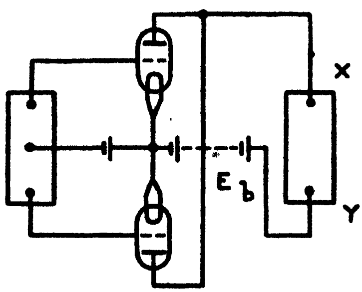


FIGURE 152.

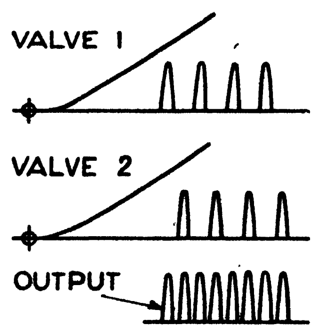


FIGURE 153.

is preferred, as the first involves a high feed and is therefore undesirable.

Beat Rectification. The special case of continuous wave detection or frequency changing by interference can be carried out by either of the above methods if the heterodyne is added to the input *AB* with the signal. A more efficient method of rectifying C.W. presents itself, however, which involves the use of the common input to the two valves.

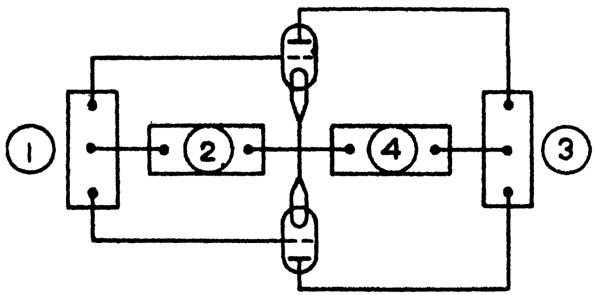
We have, as shown in Fig. 154, a circuit in which there are two inputs, 1 and 2, and two outputs, 3 and 4, all common to the two valves.

Input 1 acts differentially on the system, and it has been shown that this input with output 3 forms the push-pull circuit proper. From the push-pull input 1 no output appears at 4, as the current flows cancel. Now the input 2 acts

in parallel to the system, not differentially, and obviously any parallel impulse of the valves will lead to opposite currents in output 3, and additive currents in output 4. Thus input 2 gives output at 4, and this is a straightforward parallel valve circuit with which we are not concerned at the moment.

By itself any input at 2 is unable to influence the differential circuit 3, but if an input is applied simultaneously at 1, output will result.

For instance, consider a high-frequency signal applied to input 1, say a high-frequency C.W. telegraph dot, and a



1 — DIFFERENTIAL INPUT
3 — " OUTPUT
2 — PARALLEL INPUT
4 — " OUTPUT

FIGURE 154.

heterodyne of frequency "*f*" applied to input 2, and for the moment let the heterodyne frequency be the same as the high frequency of the incoming signal. By itself the heterodyne will have no effect on the differential output 3, but with the signal the circuit becomes unbalanced, and an output will appear at 3 provided the valves are set to rectification.

For instance, the parallel input if in phase with the differential input to valve 1, will be out of phase with input to valve 2. This means valve 1 will pass additional feed for the period of time of the dot envelope, and valve 2 passes less feed for the same period, and these effects are additive in output 3, as shown in Figs. 155 and 156. Of course a change of phase of 180° of one input would simply reverse the action, and in this

case valve 2 would have increased feed and valve 1 less, and with such phase conditions if the signal amplitude was the same as the heterodyne amplitude the feed through one valve would be proportional to double the signal amplitude and through the other zero, during the period of the dot. If the phase difference between the inputs is not 180°, output at 3 may still mature, but the amplitude resulting will not be so great.

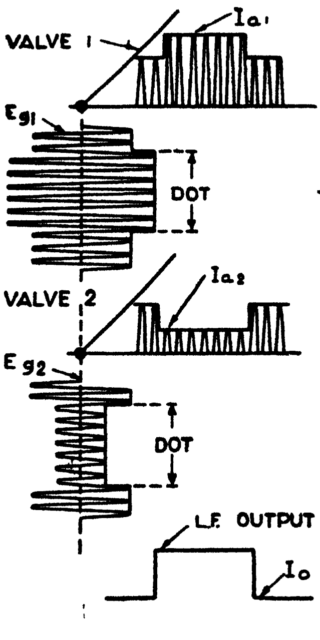


FIGURE 155.

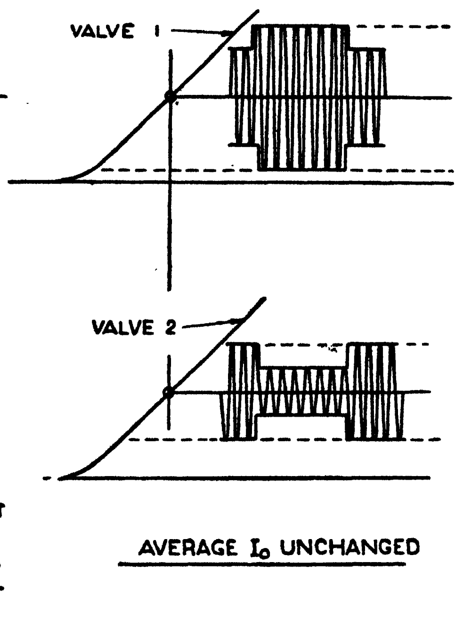


FIGURE 156.

If the two valves are set on the centre of their characteristics, there would be no signal envelope rectification at all, as the average current through each valve remains the same, varying at high frequency only, as shown in Fig. 156, but if each valve is biassed to the lower rectification point, as shown in Fig. 155, change of average feed at the signal frequency results, and the envelope can be extracted provided the circuit in output 3 is suitable.

Thus by biassing a push-pull circuit to zero, or to any point where there is asymmetry of characteristic, we can, by applying

a local heterodyne to the parallel input 2, upset the balance of the circuit and obtain the signal envelope in the push-pull output. If the local oscillator frequency is changed to produce audible beats with the incoming high frequency, the telegraph dot signal, now modulated at the beat frequency, appears in output 3 in the same manner as before. Thus for circuits where the frequency is changed, or for final detection of a C.W. telegraph signal, if a heterodyne is supplied to the parallel input 2, the incoming signal is automatically rectified, and this rectified beat may be taken out at the push-pull output in the usual manner without upsetting the balance of the circuit by direct unequal biassing.

If different frequencies (say f_1 and f_2) are applied to inputs 1 and 2, and the valves biassed to the rectifying point, frequencies as shown in the table below will be found present in outputs 3 and 4.

1	2	3	4
f	Nil	f, odd harmonics	even harmonics
nil	f	nil	f and all harmonics
f_1 and f_2	Nil	f_1, f_2 and odd harmonics	even harmonics and (f_1+f_2), (f_1-f_2)
f_2	f_1	(f_1-f_2), (f_1+f_2) f_2 and odd harmonics	f_1 and all harmonics even harmonics f_2

Special Uses of Push-Pull Circuits. Most of the foregoing discussion suggests that push-pull circuits have somewhat negative virtues, and the reader is probably at a loss to understand in what particular field push-pull is of use.

This can be explained best by outlining the various special push-pull circuits that are in common use on short waves. A first point that has already been mentioned is symmetry of circuit. For very short waves this is the most important feature of the push-pull circuit, as a symmetrical arrangement of amplifier can be balanced for zero feed-back, and balanced to earth, but this particular utility of push-pull is fully treated in the chapter on "Driven Circuits," and need not be discussed here.

It has been shown that beat rectification can be carried out, using the parallel input for the local frequency source, and

the differential input for the signal, the beat appearing in the differential output.

This circuit arrangement is of very considerable utility, both for the ordinary beat reception of a high-frequency wave and for specialised circuits such as line "tone sending," where a local tone normally suppressed in the quiescent signalling condition is keyed to line by the incoming signal upsetting the balance of the circuit.

In all cases the utility of the circuit lies in the total suppression of the local frequency source, except when the signal appears.

Ordinary high-frequency, beat-rectification provides an excellent illustration. It is very noticeable with a single

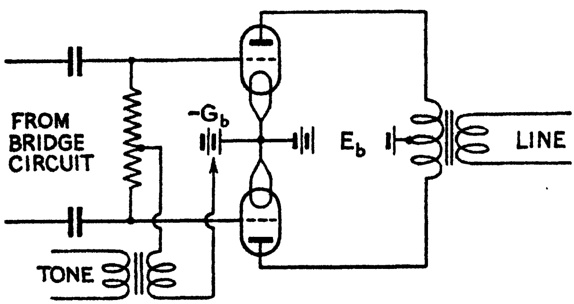


FIGURE 157.

valve rectifier and local heterodyne arranged for beat reception, that directly the heterodyne is switched on the system becomes noisy. This is due, not only to heterodyne variation, but "pick-up" on the heterodyne modulating it, this giving, not a pure effective D.C. in the output, but a modulated output.

With the push-pull arrangement, pick-up by the oscillator or variation due to change of filament, anode voltage, etc., have no effect because of the balance. Hence with this circuit the switching on of the local heterodyne does not produce a noisy condition.

A push-pull rectifier arranged for tone sending is shown in Fig. 157. The tone, whose value is arranged to suit the line, say between 500 and 1,500 cycles, is coupled to the input 2 (see Fig. 154), and on "space" no output appears. But on "mark" the circuit balance is upset, and a "mark" tone-modulated,

passed to line. The carrier-suppression circuit is very similar. The carrier is coupled to input 2 and the modulation to input 1, the side-bands only appearing in output 3.

Suppression of Second Harmonic. If the true push-pull circuit is considered, i.e. differential input and output, it is clear that any wave applied to the input is reproduced in the output as long as the total characteristic is linear.

If a large negative bias is put on the valves to cut off the feed of each valve during more than half a cycle, although harmonics are now introduced, because of the distortion of the feed current waveform, the differential action tends to produce a symmetrical wave, and thus even harmonics are eliminated. This is shown by the wave shape in Fig. 158, where "output" represents the resulting current from a push-pull circuit biassed back to a very negative value.

The importance of this circuit is found in high-frequency driven transmitters which operate with a large negative bias and give a flick impulse to the output. With single valve circuits the current wave from a single flick has a strong second harmonic, and this, if radiated, can cause interference, but the push-pull circuit automatically overcomes this undesirable feature.

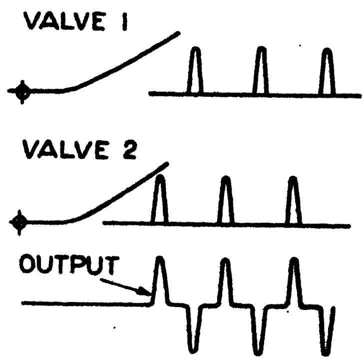


FIGURE 158.

Frequency Multiplication. The subject of frequency multiplication is a most important one in short-wave working, because it is often much easier to obtain certain effects at a lower frequency and multiply the resultant up, rather than obtain those same effects on the high frequency directly; examples are found in constant-frequency drives and in single side-band working.

In connection with this subject, one important point, which is often missed, is that to produce harmonics one does not need the original frequency-source to be rich in harmonics. In fact in most cases it is desirable that the original frequency should be fairly pure in waveform, the production of multiple

frequencies, i.e. harmonics, being solely a function of the asymmetry of circuit through which the wave is passed. A rectifier valve naturally lends itself for this purpose, and this is the simplest form of frequency multiplying device. Consider the circuit of Fig. 147, where R_a shows a triode valve backed off to any desired negative, and Z_0 an output circuit of the flywheel type, whose resonant frequency can be varied from a value f to multiples of f. Let an input pure tone of f cycles be applied to Z_I.

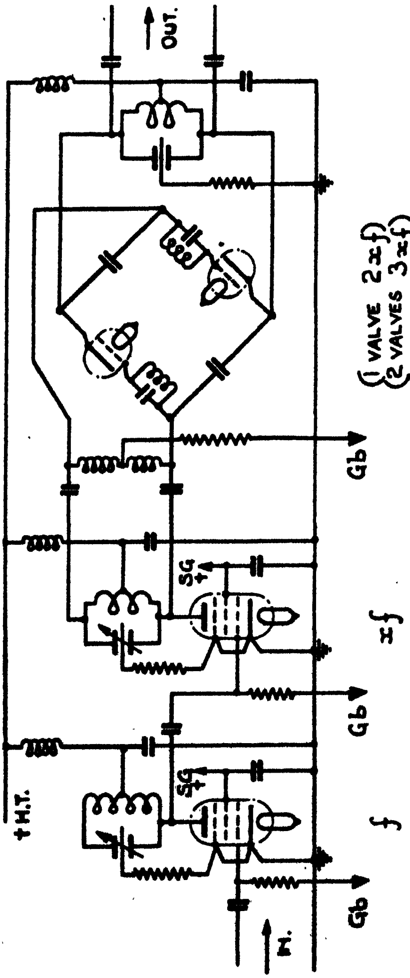


FIGURE 159.

If the valve is biassed to the cut-off point, current only flows through R_a each half cycle, and the waveform of this feed current is a series of half-sine waves. Now it is clear that such a waveform is rich in harmonics, particularly a second (twice f); the third and higher harmonics are also represented, but less strongly; hence the tuning of the circuit to the different harmonics present in the impulsing feed will lead to the output being energised in proportion dependent upon the relative strength of the harmonics in the feed, and the efficiency of the pick-up circuit.

Generally speaking, a bias rather greater than cut-off gives a strong double frequency (second) and a less strong third, the latter being increased in strength as the bias is

made more negative. Although a very large range of harmonics can be picked out from such a triode circuit the higher values are small in amplitude, even if considerable bias is used. With triode valves, it is not usual to obtain more than five times the fundamental from any one stage (however much multiplication is desired eventually), except for the purpose of certain measuring apparatus, where no appreciable power is required, and the presence of other harmonics is not detrimental. If pentodes are used a larger range of harmonics is usually possible.

For doubling purposes, a single-valve circuit is not the most efficient arrangement, because the output is only impulsed every other cycle. But if a two-valve circuit is arranged with push-pull input, and parallel output, as shown in Fig. 152, and the valves biassed to somewhat beyond cut off, a more efficient doubling arrangement results, for now the valve output circuit gets an impulse every cycle of its swing, as shown in Fig. 153. In cases where the input is allowed to run into grid current the circuit has an additional advantage over the single circuit, namely that the input is not asymmetrically loaded, this being bad because it tends to vary the original frequency-source. The arrangement just mentioned is suitable if the frequency is not too high, but with very short waves the unbalancing of the circuit by paralleling the valve anodes is not to be recommended, for, as explained previously, symmetry of circuit is the first consideration. Doubling can be obtained whilst still keeping the symmetrical push-pull circuit, by disconnecting the filament of one valve (one lead only), thus obtaining a single valve doubling circuit, the second valve acting as a capacity to preserve the bridge balance, this being explained in Chapter X. As already shown with a differential arrangement of two valves it is not possible to frequency double, except of course by unbalancing the feeds, one to saturation and one to zero, which is not desirable.

Both the methods indicated are used, the first on the earlier stages when multiplying from a low frequency such as a tuning fork; the second for very high frequencies. To obtain three times the frequency of a wave, the proper push-pull circuit can be used, with both valves operative. For if this circuit is backed off to well beyond the rectification point, the

impulsing feed is as shown in Fig. 158, and such a wave has a third harmonic very strongly represented (see Fig. 24).

Triodes, tetrodes, or pentode valves can be and are used for frequency multiplying circuits, the last named usually in the earlier stages. If biassed sufficiently negative pentodes can be employed to produce harmonics up to the 9th efficiently.

Triode circuits, which need anti-reaction balancing devices, are more normally used in the later and higher power stage, Fig. 159 showing a frequency multiplying circuit of typical design employing both triode and pentode circuits. Note, that in both cases decoupling chokes or resistances are desirable.

CHAPTER X

POWER AMPLIFIERS

VALVE transmitters may be divided into two types, self-oscillators, and driven circuits, or power-amplifiers as they are often called. The self-oscillator is but seldom used on short wavelengths except for small power transmitters, and the power amplifier is now almost universal. The general principles on which power amplifiers are built are much the same whatever the frequency, but there are, in short-wave working, several features that call for special attention.

The object of a transmitter is, of course, to produce high frequency power, modulated in accordance with the signal to be transmitted.

We desire that the power radiated should be of constant frequency and that all the modulation frequencies should be reproduced in their correct relationships. Since the power involved may be considerable, we are also interested in the power efficiency of the transmitter.

A transmitter will generally include the following features:

(1) A driving source of constant frequency. This may be of the same frequency as finally radiated or an exact fraction. In the latter case, a series of frequency-multiplying stages will be necessary and will usually form part of the master-oscillator unit proper. These multiplying stages may or may not amplify as well.

(2) A chain of amplifier stages employing triodes, tetrodes or pentodes, working at the frequency to be radiated, each succeeding stage being of increasing power.

(3) Methods for stabilising the various stages. With triodes this will take the form of some balancing system to eliminate the feedback of energy through the interelectrode capacities

of the valve. With tetrode and pentode valves such a precaution generally becomes unnecessary.

(4) The feeder circuit, coupling the final amplifier to the aerial load.

(5) The keying or modulation system with its attendant "buffer" or isolator stage.

Leaving for the moment the driving source, and the frequency-multiplying stages, both of which are dealt with in other chapters, one stage of the amplifier proper will be considered.

We will deal first with the important question of power conversion at any one stage, it being remarked that although

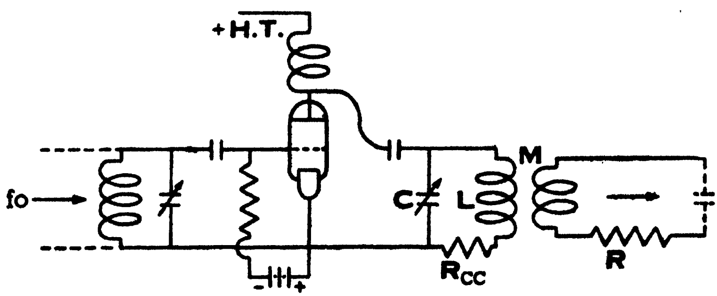


FIGURE 160.

transmitters may employ either triode, tetrode or pentode valves, from a power conversion point of view, it is largely immaterial which type is used. This will be shown later when both types of valve are discussed.

In Fig. 160 a stage of amplification is shown schematically, the problem being to deliver a given H.F. power to the output load resistance indicated by R at a frequency f_0, with least loss; R representing either the loading of a following stage or the final feeder and aerial load. The anode tuned circuit (L.C.) is usually termed a "tank" circuit because one of its functions is to store energy during parts of each cycle, so that a sinusoidal voltage is maintained across it even if the valve anode current is very distorted.

If we assume a frequency $\frac{\omega}{2\pi}$ applied to the grid of the valve and the tuning adjusted to bring the loaded tank circuit to a

unity power factor condition, then the load on the valve is an equivalent resistance of

$$R_e = \frac{\omega^2 L^2}{R_T} = \frac{L}{CR_T} = Q\,(\omega L) \quad . \quad . \quad . \quad . \quad (1)$$

(since $\omega^2 LC = 1$)

The term R_T includes the tank-circuit resistance R_{CC} and the effective resistance thrown back into the tank circuit from the output loading circuit.

Leaving for the moment the important question of the distribution of power in tank and output circuits we can see that from a power-analysis point of view we may regard the loaded tank circuit and load as replaced by an equivalent

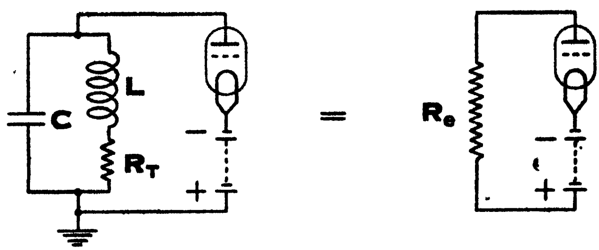


FIGURE 161.

resistance R_e as shown in Fig. 161, from which the phase relationship of anode current and volts may be observed, the grid not being considered for the present.

If a current I_a flows in the circuit we have the following conditions :

Supply volts E_b=constant.
Resistance volts $E_r = I_a R_e$.
Anode volts, $E_a = E_b - I_a R_e$.

Now, without any change of supply volts, we can change the feed current I_a by an alteration of grid volts ; further, since the load is a resistance, any change of I_a will be in phase with E_g. If, then, we increase the grid voltage in a positive direction, we increase I_a, and any increase of I_a (leading to an increase of E_r) must be accompanied by a decrease of E_a. Also, of course, the reverse will hold ; if the grid is made less positive, I_a will fall, and a decrease of I_a is accompanied by

an increase of E_a. Thus, with such a resistance load, the phases of anode and grid volts and anode feed are as follows :

Anode voltage (E_a) in phase opposition to grid voltage (E_g).
Anode current (I_a) in phase with grid voltage (E_g).
Anode current (I_a) in phase opposition to anode voltage (E_a).
Anode current (I_a) in phase with load voltage (E_{LC}).

These phase relationships will be the same whether the load is a pure resistance or a parallel-resonant circuit, with one important difference. The D.C. voltage drop in the resonant circuit being negligible, the voltage across it is purely alternating and not an alternating voltage super-imposed on a D.C. voltage as in the resistance case. Hence the voltage across the valve can rise above the D.C. value during part of the cycle, as shown in Figure 162, since the valve voltage is always the difference of supply voltage and anode load voltage. In practice the valve voltage rises to approximately twice the supply voltage.

In addition to the phase relationships in the valve itself which we have just discussed, we have also those of the oscillatory circuit currents I_L and I_C and voltage E_{LC} which are well known and are those of the ordinary parallel-resonant circuit, namely :

I_L in quadrature (nearly), lagging on E_{LC}
I_C in quadrature, leading on E_{LC}
I_a the " make up " current, in phase with E_{LC}.

These phase relationships, shown in Fig. 162, may be considered as correct as long as the driving frequency is such that the output circuit is effectively resistive in character.

To obtain maximum output, the load resistance will not require to be equal to the valve A.C. resistance, except in the special case when the valve current and voltage are sinusoidal, with no grid current and only a small amplitude of grid voltage. This is not usually a practical case, as both efficiency and output are low.

Power Output. In all high frequency power work we set the system for maximum output conditions rather than for maximum efficiency.

In order to obtain the greatest output for a given A.C. grid

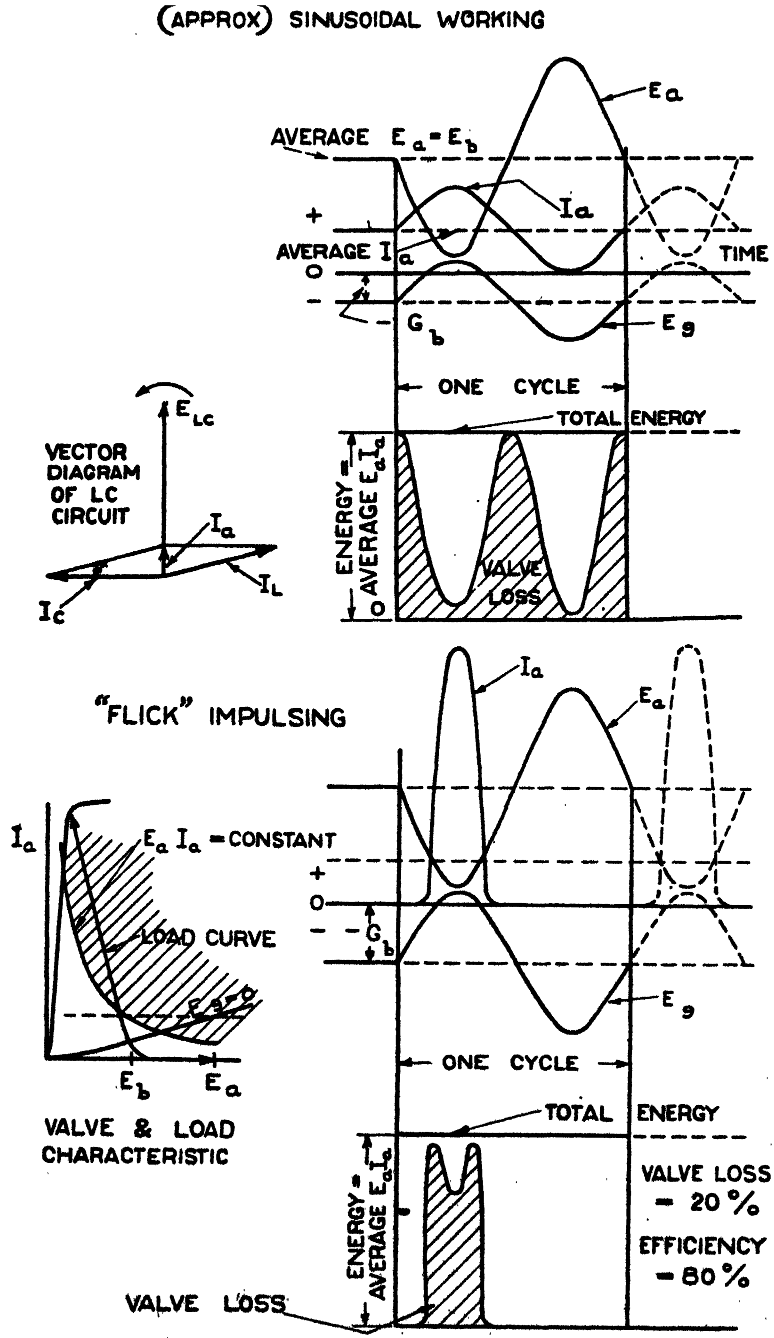

(APPROX) SINUSOIDAL WORKING
Ea
AVERAGE Ea = Eb
Ia
+
AVERAGE Ia
TIME
O
-
Gb
Eg
ONE CYCLE
TOTAL ENERGY
ENERGY = AVERAGE Ea Ia
O
VALVE LOSS
ELC
VECTOR DIAGRAM OF LC CIRCUIT
Ia
IL
IC
"FLICK" IMPULSING
Ia
Ea
+
O
-
Gb
Eg
Ia
Ea Ia = CONSTANT
LOAD CURVE
Eg = 0
Eb
Ea
VALVE & LOAD CHARACTERISTIC
ONE CYCLE
TOTAL ENERGY
ENERGY = AVERAGE Ea Ia
VALVE LOSS = 20 %
EFFICIENCY = 80 %
VALVE LOSS

FIGURE 162.

voltage and with sinusoidal anode current and voltage, the load resistance would require to be equal to the A.C. resistance of the valve. This is not usually what we require, however; in power amplifiers, and the load resistance is settled from quite other considerations and is not directly related to the valve resistance.

In practice we cannot afford to avoid the grid current region, as it would reduce our available swing of anode volts and current, and limit the useful characteristics of the valve; and hence the amplitude of A.C. voltage on the grid (usually termed "grid swing") will be made great enough to vary the anode current from zero to the filament current saturation value; further, we never work sinusoidally, but with a "flick" type of impulse, to be described later. Under these conditions, the correct load resistance to obtain maximum output is much lower than the valve A.C. resistance, and it is rather a question of utilising as much of the available voltage and current as possible.

The anode-current/anode-volts characteristic curve of a typical transmitting valve is shown in Fig. 163, where the curve for + 250 volts on the grid represents the "limiting edge" of the characteristic, so-called because curves for larger grid voltages will lie almost on top of this curve for which practically the whole electron emission from the filament reaches either the anode or grid. It will be seen that the available filament emission current in this valve is one ampere.

We will first discuss the best adjustments on the assumption that anode current is to flow during the whole cycle, that is, the alternating component of the anode current will be approximately sinusoidal.

If the one ampere anode current could be obtained when the anode volts were zero, then the load line should terminate at $i_a = 1A$, $e_a = 0$ and if the D.C. supply voltage is fixed at 6,000 volts then the loadline should be a straight line to $i_a = 0$, $e_a = 12{,}000$ volts, so that e_a varied between 0 and 12,000 volts. Such a load would utilise to the fullest extent the available emission and supply voltage. We are assuming that this adjustment is possible without exceeding the permissible anode dissipation.

The actual characteristic shows, however, that 1A cannot

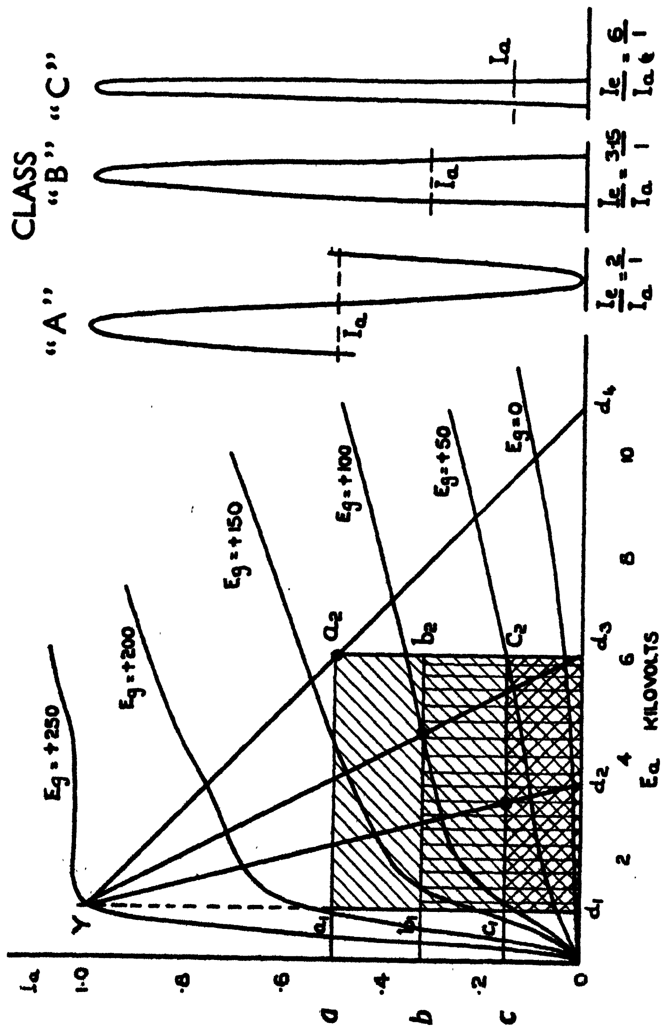

CLASS
"A"
"B"
"C"
$\frac{I_e}{I_a} = \frac{2}{1}$
$\frac{I_e}{I_a} = \frac{3.15}{1}$
$\frac{I_e}{I_a} = \frac{6}{1}$
$E_g = +250$
$E_g = +200$
$E_g = +150$
$E_g = +100$
$E_g = +50$
$E_g = 0$
E_a KILOVOLTS
E
I_a
1.0
.8
.6
.4
.2
0
2
4
6
8
10

FIGURE 163.

be obtained with less than about 1,000 volts on the anode and hence e_a can only usefully vary between 1,000 and 11,000 volts, the amplitude of the alternating voltage being 5,000 volts. The best load line is, therefore, Y, a_2, d_4, and in this case would represent an equivalent resistance $R_e = \frac{5{,}000}{0 \cdot 5} =$ 10,000 ohms. It will be seen that the mean anode current is 0·5 A and hence the input power is $6{,}000 \times 0 \cdot 5 = 3{,}000$ W and is represented by the rectangle a, a_2, d_3 0. The R.M.S. value of the alterating anode volts is $\frac{5{,}000}{\sqrt{2}}$ and of the current is $\frac{0 \cdot 5}{\sqrt{2}}$ hence the alternating power output is $\frac{1}{2} \times 5{,}000 \times 0 \cdot 5$ $= 1{,}250$ watts, represented by half the area a_1 a_2 d_3 d_1 and the efficiency is 42%.

It will be seen that the smaller the minimum anode voltage at which the full anode current can be obtained, the greater is the efficiency, the theoretical maximum being 50% when this voltage is zero. Most transmitting valves are designed so that when the maximum emission is being utilised, the minimum anode voltage will be some 10% to 20% of the D.C. supply voltage, this giving an empirical rule as a basis of design calculations both for long and short wave circuits.

These arguments apply whether a triode or a pentode valve is to be used. The only difference is in the expenditure of energy required in the grid input-circuit to reach the peak emission point. In the triode case we have to run into grid current for a considerable portion of the grid cycle, thereby requiring much power from the previous stage. This will be seen by studying the curve of Fig. 163. In the pentode case the zero grid-voltage curve is near the peak-emission curve, and hence very little grid power will be required (see Fig. 176).

We can understand easily how the power conversion will fall away with other values of R_e, for if the resistance is lowered we lose voltage swing (steeper load characteristic), whereas if it is raised we lose current swing (flatter load characteristic). With practical tuned circuits the matching to obtain maximum

output is easily accomplished, by arranging the value of $\frac{\omega^2 L^2}{R_T}$ of the tuned circuit to be correct.

Circuits should not be tapped down as the tank circuit losses are increased thereby.

Class C or "Flick" Impulsing. A different type of adjustment, by which greater efficiencies may be obtained, will now be discussed. In the case of a parallel-resonant circuit the voltage across it and the current circulating in it remain very nearly sinusoidal even if the current entering it departs very much from a sine-wave form.

As explained, the parallel-resonant circuit acts like a fly wheel and its efficiency as such will depend upon the ratio between the energy oscillating in it and the energy dissipated in it per cycle, that is, upon the ratio $\frac{kVA}{kW}$, or the Q value, since $\frac{VA}{W} = \frac{\omega LI\,(I)}{I^2 R} = Q$. The smaller this ratio the greater the efficiency of transfer from tank to output but the greater the harmonic content.

A special case of "flick" impulsing which we will term "Class B" is that in which the anode current is flowing for 180° of each cycle and consists therefore of half-sine waves. The load line will then be $Y\,d_3\,d_4$. The equation for a series of half-sine waves is:

$$i_a = \frac{I_{max}}{2}\left[\frac{2}{\pi} + Sin\ \theta - \frac{4}{\pi}\cdot\frac{1}{3}\ Cos\ 2\theta\ \ldots\ldots\ldots\right]$$

The first term is the D.C. component and, when multiplied by the supply voltage, gives the input power. The average current is now $\frac{I_{max}}{\pi} = \cdot 318$ amps., and the rectangle $0\,b\,b_3\,d_3$ shows the input power. The alternating voltage is, as we have already explained, very nearly sinusoidal and hence the output power is given by the product of R.M.S. voltage E, by the R.M.S. fundamental current. It will be seen from the series above that the fundamental current is:

$\frac{I_{max}}{2}\ Sin\ \theta$, this having a maximum value $= \frac{I_{max}}{2}$.

Thus the power output $= \frac{1}{2} \cdot \frac{I_{max}}{2} E_{A.C.\,max}.$

The remaining terms in the expression for i_a will convey no power because there is no corresponding harmonic in the voltage wave. The efficiency is improved to 66% because the anode current now flows mainly when the anode voltage is low so that the product of anode current and anode voltage integrated over a cycle (the valve anode loss) is smaller.

If the grid bias be still further increased, then the anode current will flow for less than half of each cycle and this we may term a "Class C" type of impulse, and it is specified by its angle of current flow. The case in which the angle of current flow is 90° is shown in Fig. 163, the load line being $Yd_2 d_4$, the average current $=\cdot 163$, and the input represented by the rectangle $Oc c_2 d_3$. Actually "Class C" operation is usually arranged so that angles of current flow are between 120° and 150°. A greater angle is not used as the efficiency is lower, and a lesser angle necessitates very large grid-bias and grid-swing voltages. It should be pointed out that both with Class B and Class C working the slope of the load line will no longer be that of an equivalent resistance $\frac{\omega^2 L^2}{R_T}$.

It is clear that for the three types of adjustment, input, output, and efficiency vary widely and the results are tabulated below.

TABLE I.

D.C. Supply Voltage 6000 V. Peak Emission 1A.

Minimum Anode Voltage 1000 V. Alternating voltage (max. value) 5000 V.

Class.	Anode Current.		Ratio Peak/ Mean.	Input Power Watts.	Anode Loss.	Output Power Watts.	Efficiency %
	D.C. Compt.	Fundamental A.C. Compt. (max. value).					
A	0·5	0·5	2	3000	1750	1250	42
B	0·318	0·5	3·15	1890	640	1250	66
C (½ sine wave.)	0·163	0·310	6·15	978	203	775	79

Thus for the same peak emission it will be seen that the efficiency increases and that both input and output decrease as the duration of the current pulse is reduced. It will be necessary to take into consideration the permissible anode dissipation and hence the greatest output may be possible with Class C adjustments, because of the increased efficiency. The general statement sometimes made that the sinusoidal (Class A) adjustments give the greatest output is, therefore, frequently not true in practice.

It will be evident that a valve to be used under Class C conditions should have a filament capable of giving an emission current several times the mean anode current which can be passed to the anode without over-heating it. Generally speaking, the filaments of valves designed for Class C adjustments provide a peak to average emission of between 6/1 and 4/1, the anode being capable of dissipating 25% of the power at the average I_a E_a value. It will therefore not be possible to plot the static characteristic of such a valve or run it at full output under sinusoidal conditions.

Under the conditions obtaining with Class C working, the valve is not to be regarded as a high frequency alternator at all, but rather as a commutator making and breaking a direct current supply at the frequency of the L.C. circuit.

A perfect commutator is a power converter of D.C. to A.C. (square waveform) and has the same phase relationship of current and voltage as the valve, but since the volts drop to zero when full current passes and the current to zero when it has the supply voltage across it, the efficiency as a converter is 100%. In the case of the valve, however, the limiting edge shows that at the time of passing maximum current the volts have a finite positive value (some 12% to 20% of the total D.C.), the change of state is not instantaneous, and under the best conditions the power conversion is never greater than 85%.

Design of Tank Circuit. We have shown that the conversion efficiency from D.C. to H.F. is determined by the type of anode-current wave used, and on the correct impedance of the loaded tank circuit to utilise the peak emission of the valve under the conditions appertaining, i.e. whether Class B or Class C. It is, however, equally important to design the tank so that its losses are the least possible.

If the output is coupled through a mutual inductance M, and the load circuit is in resonance, we have :

$$w\, M\, I_{\infty} = R_o\, I_o.$$

The power in the load is $I_o^2\, R_1$ or in terms of primary current, $\frac{\omega^2 M^2}{R_o} I_{\infty}^2$ and the power input to the tank circuit is

$$I_{\infty}^2 \left(R_{\infty} + \frac{\omega^2 M^2}{R_o}\right)$$

The ratio $\frac{\text{power transferred to load}}{\text{power input to tank circuit}}$ may be termed the transfer efficiency of the tank circuit and is evidently given by

$$\frac{\dfrac{\omega^2 M^2}{R_o}}{R_{\infty} + \dfrac{\omega^2 M^2}{R_o}}.$$

If Q_1 and Q_2 are the "Q values" of the tank circuit when unloaded and loaded respectively, then $Q_1 = \frac{\omega L}{R_{\infty}}$ and $Q_2 = \frac{\omega L}{R_{\infty} + \frac{\omega^2 M^2}{R_o}}$. Hence $\frac{\omega^2 M^2}{R_o} = \omega L \left(\frac{1}{Q_2} - \frac{1}{Q_1}\right)$ and transfer efficiency may be written as

$$Q_2\left(\frac{1}{Q_2} - \frac{1}{Q_1}\right) \text{ or } \frac{Q_1 - Q_2}{Q_1} \times 100, \text{ as a percentage.}$$

From the above it is seen that the greater the difference between Q_1 and Q_2 the greater the transfer efficiency. Assuming the unloaded tank circuit Q_1 is fixed irrespective of its L.C. ratio (a justifiable assumption), although Q_2 should be made as low as possible to obtain the highest transfer efficiency, in practice the minimum value of Q_2 depends upon a number of factors.

The lower we make Q_2, the greater the harmonic content as previously explained, and the less the coincidence between maximum output and unity power factor. On the other hand, an increase of Q_2 reduces the transfer efficiency and if increased too much (see page 6) may introduce "sideband cutting." Actual values of Q_2 range from 15 to 20 on small

power sets down to 3 and 4 on larger sets, and transfer efficiencies become possible up to values as high as 95%. Although it is not obvious from the above formulæ, we obtain the greatest transfer efficiency by increasing the $\frac{L}{C}$ ratio to the maximum possible, but here again the type of circuit will probably determine the actual design and $\frac{L}{C}$ used.

For instance, consider a triode transmitter designed to give an output of 1 kW on a minimum wavelength of 15 metres from a D.C. supply voltage of 5,000. The valve and stray capacitances of the output circuit (including that of a small tuning capacitance set to minimum) to deliver this power could easily be as much as $20\mu\mu F$. Thus with no additional deliberate tuning capacitance we should need an inductance of only $3{\cdot}16\ \mu H$ to tune to 15 metres.

If we assume the minimum anode voltage is 1,000, then the R.M.S. value of the A.C. anode voltage is 2,830 volts and we have $I_{ac} = E\omega C = 7{\cdot}13$ amps.

$$\text{Now } Q_2 = \frac{VA}{\text{watts}} = \frac{7{\cdot}13 \times 2{,}830}{1{,}000} = 20{\cdot}2.$$

This is the minimum Q_2 value that can be obtained with a circuit layout and valve having this particular stray capacitance, operating on the voltage mentioned.

It is observed also that if the valve capacitance forms a considerable part of the total capacitance, as it would do in such a case, the valve carries an equivalent proportion of this total tank-circuit current, and therefore the valve electrode seals need to be designed to carry the high-frequency current in addition to the feed. In glass-envelope valves special precautions have also to be taken to prevent "hot spots" forming on the envelope due to eddy currents produced in any accidental metallic deposits produced in the manufacture of the glass. On early forms of valves, metal strip jackets are desirable to equalise the surface potentials, but in modern valves any deposits formed during manufacture are washed from the envelope walls.

In addition to keeping up the $\frac{L}{C}$ ratio of the tank circuit,

it is also necessary to use the minimum amount of insulating material, and where insulators are used they should, as far as possible, be kept outside the high frequency fields, suitable types of material being dealt with later in the chapter.

If an amplifier stage is set up with triode valves, as shown in the schematic diagram of Fig. 160, it will be found to oscillate due to the interelectrode valve capacity C_{ag} forming a reversible coupling which can cause feed back. Hence before such a circuit can be driven it is necessary to eliminate the tendencies to oscillate by providing anti-reaction arrangements.

Triode Anti-Reaction Circuits. The simplest arrangement is as shown in Fig. 164a (diagrammatically in Fig. 164b).

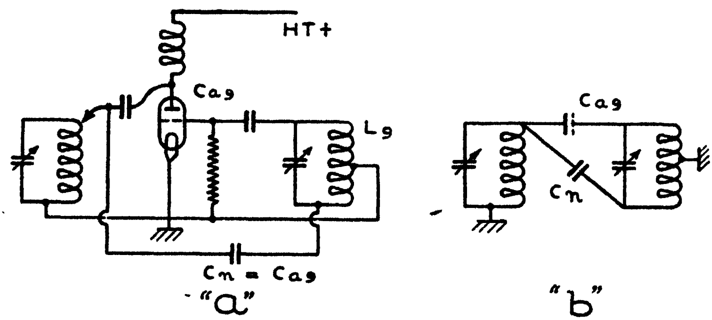


FIGURE 164.

Here the input coil is split and the filament tapping taken to the centre so that grid/filament is across but half the input coil L_g. One end of the coil is already coupled to the output circuit through the grid anode capacity of the valve C_{ag}, and hence if the other end of the coil is connected through the condenser C_n to the anode it will act as an anti-reaction condenser, and balance the grid/anode capacity coupling. Thus power in the output changes the potential of the anode relative to the filament and hence will create currents through the condensers C_{ag} and C_n which flow to earth through the input coil in opposite phase, and so cause no feed-back voltage between grid and filament.

An alternative arrangement is shown in Fig. 165a (diagrammatically in Fig. 165b). In this case the whole input is left across grid-filament, but the output centre-tapped instead, and the free end connected back to the grid. Since the anode

supply will be connected between filament and the tapping point on the coil, the circuit is now at high potential, and this is undesirable on high voltage systems. Otherwise the action is similar to the previous arrangement. It is a matter of convenience which method is adopted and both are found in short wave circuit practice, but on very short waves and large powers this simple balanced arrangement is not too satisfactory.

Examination of the above circuits shows that although the valve capacity has been neutralised, neither circuit is symmetrical as regards earth and any asymmetry of circuit is undesirable. In fact, ordinary methods of balancing are of doubtful value if high efficiency is required, because valve

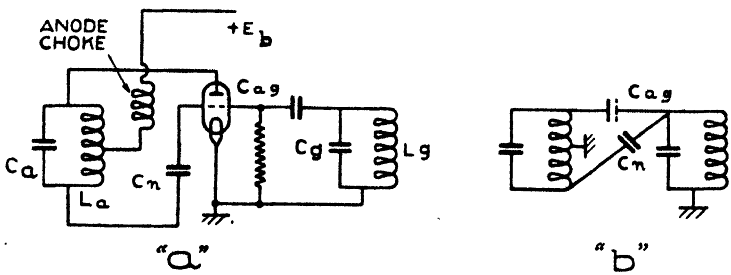


FIGURE 165.

capacity is not the sole factor to contend with, circuit layout is more important.

If we build a geometrically symmetrical circuit, with earth as a datum, we have the best chance of obtaining stable results on very short waves, provided all connecting leads are reduced to a minimum; which is another way of stating that in balancing a circuit to achieve zero feed-back, not only must all obvious coupling be balanced but capacity effects to earth must not be forgotten. As our minds are not accustomed to think in terms of the minute values which are only significant at these very high frequencies, the mechanically balanced layout automatically helps us to obtain the desired results.

The Bridge Balance. C. S. Franklin developed a truly balanced system by building a circuit from a bridge point of view, that is to say taking the valve capacity C_{ag} as one arm of a bridge, three other capacities are provided to complete

the bridge and the whole circuit built to fit this symmetrically, not only as regards input and output but with due regard to the reference point earth, so that no coupling exists between circuits placed across the diagonals of the bridge arms, as shown in Figs. 166a and 166b. On wavelengths not too short and small powers, the circuit as shown will give a clean balance; but on waves below some 25 metres, the circuit in the simplest form described above is not perfect, owing to the fact that the bridge shown is but a pure capacity balance and no account has been taken either of the power factor of the condensers making it up or the resistance of the conductors. Since the valve is a leaky condenser, it is necessary to add resistance to the

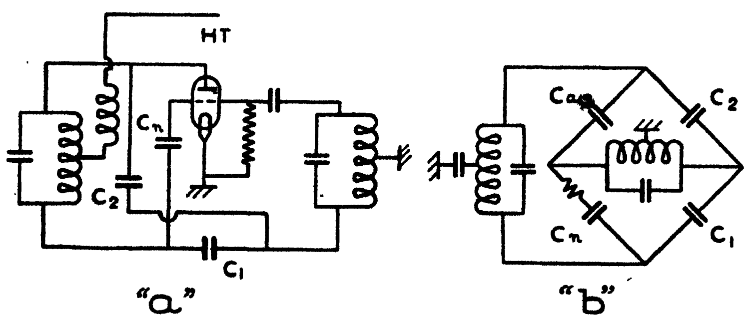


FIGURE 166.

complementary bridge arm to compensate for this, and either a large shunt resistance or a small series resistance can be used. It is simpler to use a series resistance, as only a fraction of an ohm is required, and with this added compensation a perfect bridge balance can be obtained.

It is but a step to turn the single valve bridge circuit to a two-valve bridge for the purpose of handling more power, the condenser C_1 being replaced by a second valve, as shown in Figs. 167a, 167b, and this circuit will now be recognised as a push-pull type, symmetrically balanced for anti-reaction.

The layout of the bridge circuit must be such that leads to the grids of the valves, and in fact leads generally are made as short as possible. On very short wavelengths, however, the length of lead into the valve may be sufficient to upset the balance.

For although the corners of the bridge arms are in opposite phase the reactance of the leads from the corners to the valves throws out the anti-phase voltage condition on the grids. To compensate for this, condensers are necessary in each lead whose reactance will cancel the lead reactance. These condensers C_3 and C_4 are shown in Fig. 167, the chokes L_3 and L_4 being to provide the necessary D.C. current path and prevent grid blocking.

The number of amplifier stages required depends not only on the characteristics of the valves used and the type of the circuits, but on the power of the driving source and the final power output required. With a well-balanced bridge the

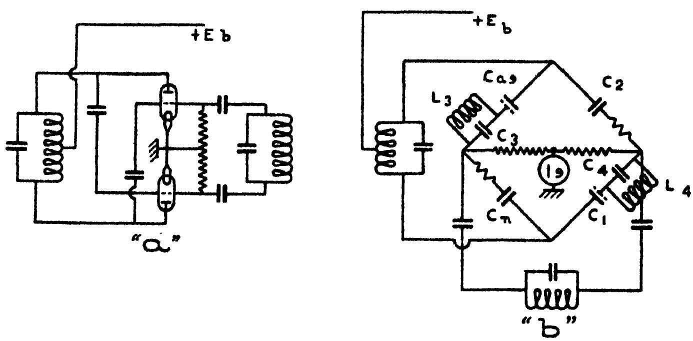


FIGURE 167.

step up in power ratio may be as great as 60/1 on wavelengths as low as 20 metres, although, generally speaking, a normal ratio is 25/1.

Fig. 168 shows a three-stage amplifier system which depicts both single and two-valve bridges, the driving source and keying arrangements being omitted.

Experimental Analysis of Driven Stage. It is of interest to make an analysis of a triode driven stage and consider the means adopted for adjusting it.

When thinking of self-oscillators and driven transmitters, there is a tendency to class them together as being very similar in action, whereas, except for the fact that our aim is the same in each case—to produce high frequency power—there is considerable difference in the method of obtaining this result.

The main difference in action between the oscillator and the amplifier may be summed up as under :

1. (*a*) A self-oscillator always sets itself to that frequency where the total reactance of the system is zero. This means one can always consider the load on the valve as resistive, or almost so.

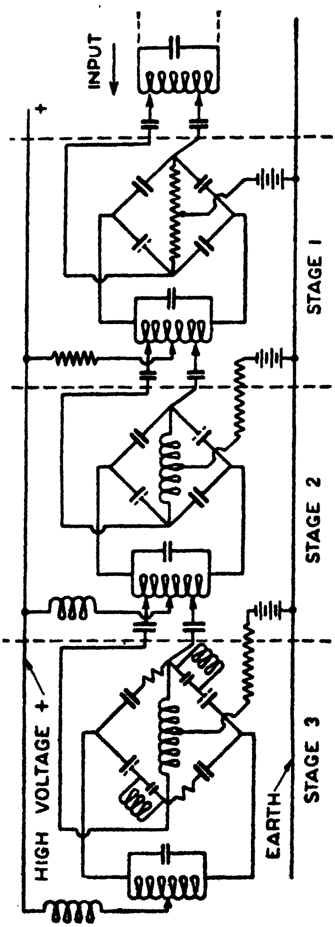


FIGURE 168.

(*b*) In a driven stage, since the grid input frequency is fixed by an independent circuit, the tuning of the output will vary the load reactance and hence the driven or amplifier valve will only operate at "unity" power factor at one setting, namely when the natural frequency of output is that of the driving E.M.F.

2. (*a*) In a self-oscillator the grid swing is a function of output current, and hence the former is automatically limited when optimum conditions are reached. For any move to force increase of swing is checked by falling away of output.

(*b*) In a driven stage the grid swing is to some extent independent of the amplifier condition, and hence one can drive the amplifier grid to a voltage in excess of, or less than that required to give optimum conditions.

3. (*a*) The wavelength of the grid circuit in the case of the self-oscillator needs to be away from that of the output for efficient conditions.

(*b*) In a driven stage the grid wavelength must be the same as the output.

4. (*a*) For coupled circuit working the limitation of output from a self-oscillator may be controlled by an instability effect. Hence, in such cases a self-oscillator cannot be loaded up to its full capacity.

(*b*) In the driven case this is not so, and coupling to a low damped secondary can be made to the full amount essential for the complete loading of the amplifier.

5. (*a*) The frequency stability of a self oscillator is dependent upon the flywheel effect of the tank circuit. This necessitates a fairly large ratio of kVA/kW for the output, usually in excess of 20/1.

(*b*) In a driven circuit we only require the flywheel effect to reduce harmonics, and in consequence the ratio kVA/kW is made small in order to get the best transfer efficiency.

For the purpose of analysis, a system with a separately tuned grid circuit has been chosen, with the valve anti-coupled by means of a condenser. The full bridge has not been adopted in order to simplify the plotting of characteristics, and the circuit analysed therefore is similar to that shown in Fig. 165, with a drive applied by mutual coupling to the coil L_g. The wavelength is 28 metres.

To commence with, the circuits L_aC_a, L_gC_g must be tuned to the driving frequency, and the value of the anti-reaction condenser must be found. These adjustments must be made simultaneously, as the alteration of one affects the other. The most efficient method is to drive at low power and find the point of zero output, using a sensitive meter in the amplifier circuit L_aC_a, with the amplifier valve filament on and with, or without, the lower end of the grid leak connected; but with no voltage on the amplifier anode, of course.

The filament is kept on, as the valve capacity is more nearly that obtaining under working conditions; and with the grid leak connected, grid current is obtained which assists to tune the circuit L_gC_g. But the final balance should be found with the grid leak disconnected in order to remove grid current damping and sharpen up the tuning of the circuit L_aC_a, the two curves of Fig. 169 showing the change of output and grid current for different positions of anti-

balance condenser. Continual retuning of the amplifier circuits is necessary as the balance point is approached, and the circuit L_sC_s is finally tuned as near the balance as possible.

Observe that the point of balance is given by minimum current appearing in the amplifier output, for at this point the circuits are uncoupled; or by maximum grid current in

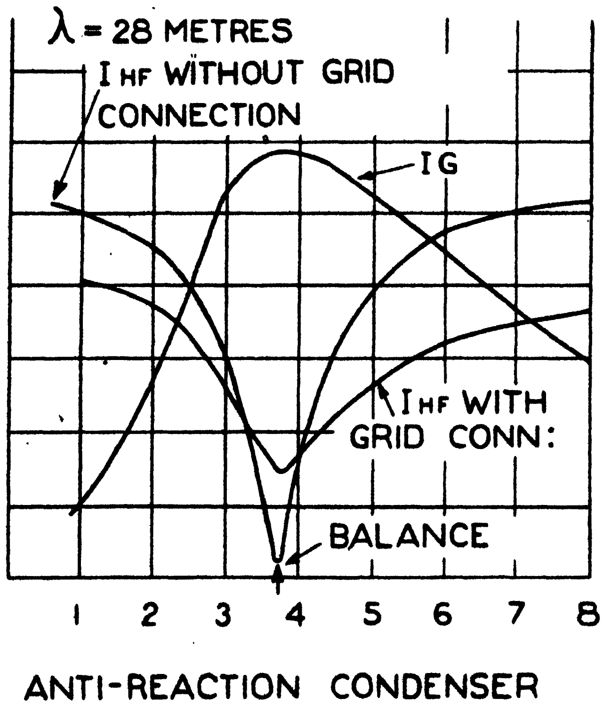


FIGURE 169.

the case where the grid lead is left connected, for at the balance point there is no coupling to the amplifier output and hence the load on the grid circuit is least.

We will now consider the driving conditions with power on the amplifier.

The frequency of input voltage being constant, the type of load experienced by the amplifier will depend on the tuning of amplifier output circuit relative to the input frequency, and only at one point will this output circuit be non-reactive in

character. Hence the load line characteristic will vary from a straight line to an ellipse, as the output circuit is detuned.

If the tuning of the output is varied, the load on the valve becomes reactive in character and of smaller value, because away from resonance the impedance of a parallel LC circuit falls. The effect of this is observed by a rise of anode feed and by greater loss in the valve and consequent loss of efficiency. The feed rises because of the lower load impedance and

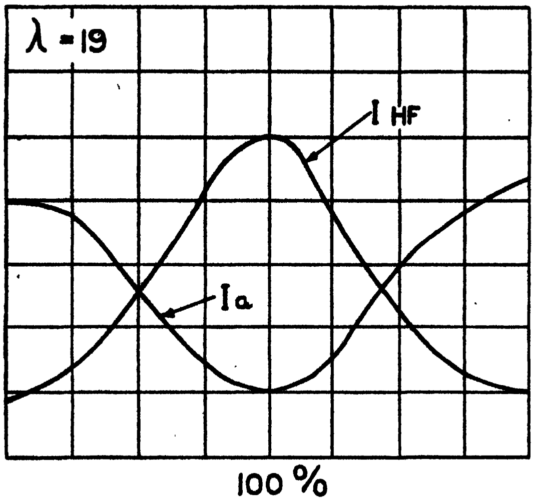


FIGURE 170.

the valve losses increase because the valve current is no longer in phase opposition to the valve voltage.

If the balancing of the amplifier stage has been made carefully as detailed in the last section, then the application of supply voltage to the anode of the amplifier should result in the correct resistive load condition being obtained, and if so the feed taken by the amplifier will be the minimum possible. In fact the final setting of every amplifier is accomplished under working conditions of full output by adjustment of output tune to give *minimum* feed, a tuning either side of this leading to increased feed and greater valve loss.

This effect is shown in the curves of Fig. 170, where minimum

feed indicates the resistive load condition and point of minimum valve loss. These curves were taken by varying the tuning of the amplifier output circuit, the anti-reaction position being fixed to the correct setting, as indicated by minimum current in the curve of Fig. 169. The point of minimum feed, which will always be the point of maximum efficiency, may or may not coincide with maximum output from the amplifier, but it will coincide if the anti-reaction condition is perfect, and if the Q value of the output circuit is not too low.

Input Voltage Required. So far nothing has been said of the power required to drive an amplifier stage, but this is of the greatest importance, for under-driving leads to a reduction of efficiency and output, and over-driving involves an unnecessarily large preceding stage, although it is better to over-drive a telegraph transmitter rather than under-drive. This question can best be studied by examination of the curves of Fig. 171, which shows that the efficiency rises from zero to about 75% step by step with increase of power output. Thus under-driving leads to low efficiency and low power output.

These curves, which were obtained with the circuit of Fig. 165, show, of course, only the amplifier power conversion and do not take into account the input power from the previous stage. A somewhat noticeable point is that it does not appear to be possible to over-drive an amplifier stage for the output and efficiency are seen to remain high with very large values of input E.M.F. This is so, and the explanation is to be found in the grid current, which is observed to increase rapidly after maximum driving conditions are reached, further driving voltage to the grid increasing grid current; the increased damping is such that the actual voltage measured across L_g,C_g does not increase; since it is this voltage which swings the amplifier grid, the result is merely a reduction of the total overall efficiency because the input power is unnecessarily large.

This result is the same whether leak or battery bias is used, although in the former case the effect is less noticeable, on account of the grid current increase forcing back the static negative grid bias, and so preventing the grid current rising so rapidly.

The foregoing discussion and methods of adjustments may be considered as applying to every type of driven amplifier, and where one has a chain of them, the systematic procedure of adjustment indicated is most essential if the best results are desired.

Filament Emission. A transmitting valve is very sensitive to filament emission limitation, and if this emission is reduced below a certain value no output at all can be obtained, the

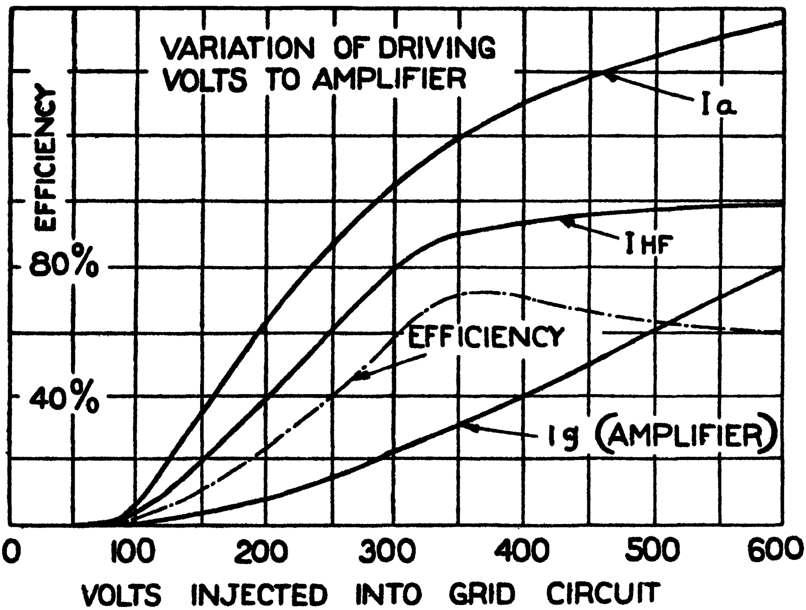


FIGURE 171.

curves of Fig. 172 showing the effects resulting from reduction of filament voltage.

It is observed that below a certain critical value, indicated by a rapid fall of grid current, the output falls away rapidly, and this is due to secondary emission taking place in the valve.

It is a well-known effect on long waves, but appears to be much more marked on short waves, and insufficient filament emission is the cause of many transmitters not functioning.

Parasitic Oscillations in Triode Amplifier Stage. In a complex circuit it is evident there are many possible

"degrees of freedom," and where amplifying valves are employed spurious forms of oscillation are often in evidence. These unwanted oscillations manifest themselves chiefly in multi-valve transmitters on large powers, and are worse with valves having a high mutual conductance.

Parasitic oscillations show up in a variety of ways: by high anode feed currents, grid currents of unusual value, circuit

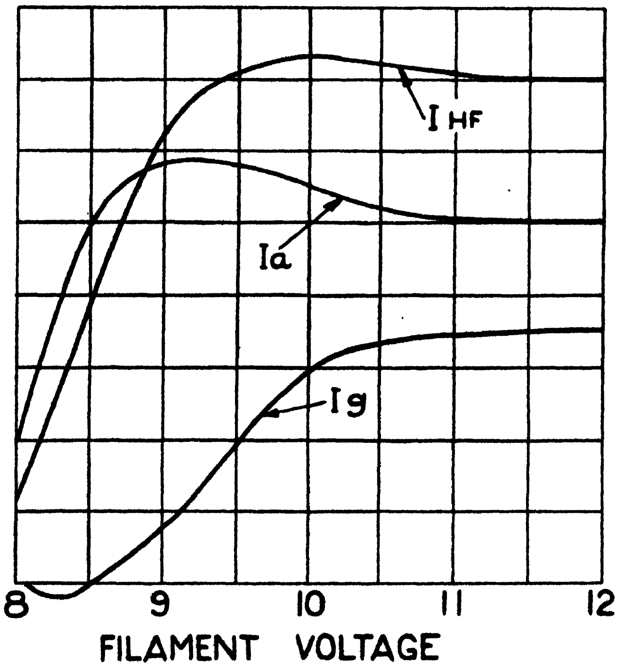


FIGURE 172.

instability and violent voltage transients; and sometimes by an audible note when listening to the radiated wave in a tuned receiver. The ordinary anti-reaction arrangements do not stop these spurious oscillations, and as the effects are altered considerably by slight differences of circuit layout the exact arrangement to prevent a particular parasite is not calculable beforehand, but must be developed to meet each case.

It is difficult to classify parasitic oscillations, but there are one or two fairly well-defined types which may be mentioned.

(1) **Parasites of a Very Short Wavelength.** The inductance of connecting leads between valves and circuits, together with the valves capacities, may form a circuit of very short natural wavelength. Such a circuit may be found connected, not only with the high frequency valves of a transmitter, but also with the modulators, and in the event of self-oscillation, considerable power may be developed in the parasite.

When the parasite occurs in long-wave, high-frequency circuits or in modulating circuits, it can be prevented by providing a path of high resistance in the grid-anode valve leads, and a low impedance path from grid to earth at the parasitic frequency. This is accomplished by the addition of a resistance, usually of value between 50 and 500 ohms, in the grid or anode circuit, and a condenser shunt from grid to filament.

On short wavelengths such arrangements are not possible, as they will also prevent efficient operation on the desired wave, and hence this particular form of parasite must be avoided by careful design and layout of the circuit, such that leads from valves to circuits are as short as possible or are compensated for.

(2) **Parasites near the Fundamental.** This parasite may cause audible modulation and in some cases is due to a choke, or choke condenser combination, or on short waves even a lead with its self-capacity to frame, having a natural frequency too near the fundamental and producing self-oscillation at a particular frequency, which beats with the fundamental.

If an amplifier circuit is not properly balanced, or direct coupling exists between stages, somewhat similar effects may be observed.

The presence of audible modulation, or howl, may also be brought about by a high resistance in the grid circuit of a stage which is self-oscillating. This effect, which was called originally a "squegger," is well known and need not be discussed.

Inverted Amplifier. An original amplifier circuit devised by Standard Telephones and Cables Ltd., and used by them on the final stages of their short-wave sets at the B.B.C. Station at Daventry is "series connected" or "inverted," as it is called. A simplified diagram of connections is shown in Fig. 173, and an equivalent circuit in Fig. 174.

The unusual feature is seen to be that on the input side, the grids are earthed and the filaments are at the high-frequency potential. In the actual amplifier the grids are earthed through series condensers which neutralise the inductive

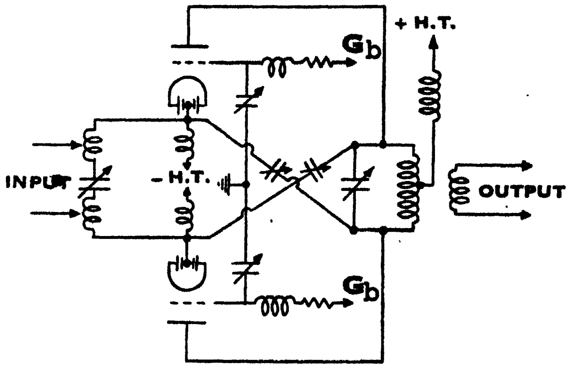


FIGURE 173.

reactance of the grid leads and these condensers must therefore be adjusted for each change in operating frequency.

It will be seen that if the grid leads are effectively earthed to high frequency then they form a screen between input and output circuits similar to the screen in a tetrode or pentrode valve. The tendency to self-oscillation is thereby reduced and when neutralising condensers are necessary they will be

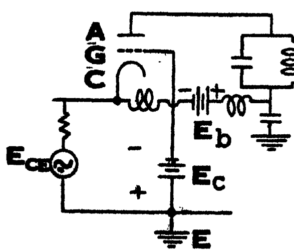


FIGURE 174.

very much smaller than in the conventional triode circuit, which assists in the design of high-power, short-wave amplifiers. The capacity across the anode tank circuit is also less with the inverted arrangement because of the small size of balancing condenser.

In Fig. 174 one half of the push-pull arrangement is drawn in its simplest form where A, G, C are the anode, grid and cathode of the amplifier valve, and in Fig. 175 is shown the corresponding curves for one cycle of potential between them. E_{CE} is the voltage between cathode and earth from the exciter stage, E_{GC} the voltage between grid and cathode of the amplifier, and E_{AC} the amplifier anode voltage, from which it is seen that the amplifier and exciter voltages are in phase, but as usual the amplifier grid and anode voltages are in anti-phase. Further that the amplifier valve and exciter circuit are in series across the output circuit and the phase relationships are such that the drive supplies some power to the output stage. The exciter circuit has, therefore, to be of larger

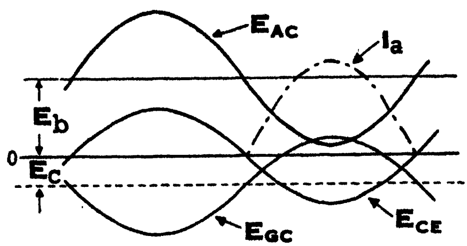


FIGURE 175.

rating than in the conventional design but some of the extra power is utilised in the output.

The alternating component of the amplifier anode current evidently flows through the impedance of the exciter circuit and this provides negative feed-back of an amount which depends upon the adjustment of the exciter circuit.

When such an arrangement is modulated it is necessary to modulate the exciter stage as well as the amplifier because of the fact that some of the drive power appears in the final output.

Pentode Amplifiers. The triode has for many years been, and still is, a most efficient type of valve for H.F. transmission work. Special designs of triode valves have been evolved (see Appendix V) to meet the particular requirements of high and ultra-high frequencies; and conversion efficiencies can be obtained with a modern S.W. triode as high, or even higher, than normally obtained at lower frequencies.

Beam-tetrode and pentode valves for transmitting work are, however, also in common use in the medium and lower power sets. The beam-tetrode is a screen-grid valve designed so that secondary emission from the anode is eliminated by the concentration of the electrons into a beam, which increases the space-charge effects existing in the anode/screen space and ensures that a potential minimum is created near the anode surface even at low anode-voltages. This effect is achieved by the aligning of the grids in the valve and the inclusion of two beam-forming electrodes, parallel to the screen supports and placed between the screen and anode. Such valves are not usually more efficient than the triode, but where flexibility is required they have certain advantages because their low inter-electrode capacity eliminates the use of neutralisation arrangements and this simplifies the circuit layout. In the case of the pentode valve, it is possible to carry out effective low-level, 100% modulation by using the suppressor-grid as a control electrode.

The design of beam-tetrode and pentode valves for transmitting work calls for features which are somewhat different from those met with on similar valves developed for receiving work because the avoidance of power loss is an important factor, and the seals of the various electrodes will be required to carry considerable high-frequency currents, and must be designed to have the minimum possible resistance and inductance. In a modern transmitting pentode (see Appendix V), a conventional seal will be used for the control electrode g_1, but the connecting wires of the screen-grid g_2 and the suppressor-grid g_3 are of multiple form and are taken away through the base of the valve. By this means, the inductance of the leads is reduced to the minimum possible, the current distribution is made uniform and the current-carrying capacity large.

One of the advantages of the pentode valve is the ability to use suppressor-grid modulation (see page 419) and this necessitates certain guiding rules in designing the valve. If its amplification factor μg_3 is made high, then only a few volts swing are necessary to control the anode current fully, but the peak anode current can only be reached with positive volts on g_3. This is an undesirable feature for modulation

purposes, as grid current will be taken, which would load and distort the modulation output. If the amplification-factor is made too low, the peak anode-current can now be obtained with zero or even negative voltage on g_3, but the output is lower than with the previous design and a considerable voltage swing on g_3 is required to fully modulate the anode current. A further consideration is that the design of g_3 also affects the screen (or g_2) current, the higher the amplification of g_3 the less the screen current, and because only small changes of g_3 will fully modulate the valve, these changes will be accompanied by only a small change of screen current. Thus when the valve is designed for suppressor grid modulation a compromise is adopted, the peak anode current being obtained with g_3 at zero voltage or very slightly positive.

Another important feature in the design of both the beam-tetrode and pentode valves is the reduction of the screen losses to the minimum possible. As has been mentioned, this can be helped by keeping up the amplification-factor μg_3, but mostly it is brought about by the principle of alignment, i.e. the grid structures are not wound random fashion, but correctly aligned one with the other so that each g_2 wire is exactly behind a g_1 wire, and similarly with each g_3 wire. The characteristics of a typical pentode transmitting valve are shown in Fig. 176, from which it is observed that the working screen-potential in this particular case is some 20% of the anode-voltage, and under these conditions it is possible to obtain peak anode emission when the anode potential is reduced to some 15% of the working average potential, i.e. giving similar conditions to those obtained with a triode valve. Examination of these characteristics will reveal that if one operates such a valve under the condition say of Class B working, as shown by the dynamic characteristic in Fig. 176, the conversion efficiency from D.C. to H.F. will be similar to that obtained with a triode.

Overall Efficiency of Pentode Valve. We will now compare the efficiency of the pentode valve with that of the triode. In considering the true efficiency of any form of amplifier, we need to consider power in the cathode circuit and also losses in any grid circuits as well as conversion efficiency of the anode circuit. In the triode case we have

only grid g_1 loss, but in the pentode case we have g_1 and g_2 loss to be considered. In the pentode valve the loss in g_1 is much less than in the corresponding triode valve, because

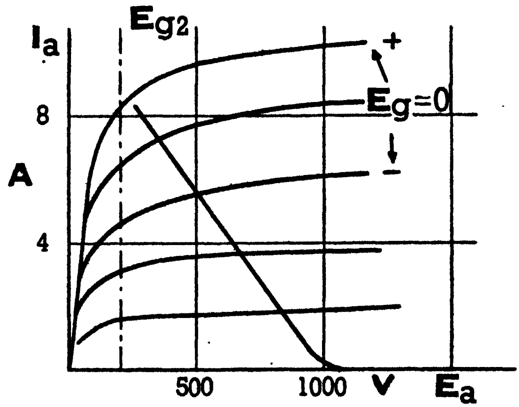


FIGURE 176.

excursion into the grid current region is much less in the case of the pentode than in the case of the triode. We must also remember that the g_1 loss is an H.F. loss, and it is important to reduce this as much as possible as it has to be supplied from a previous stage which is itself working at a conversion efficiency of perhaps 70%.

In the case of g_2, or screen loss, this is, of course, absent in the triode valve. In a well-designed pentode the screen voltage is as low as 20% of the anode voltage and the screen loss is twofold. First, there is the actual loss in the valve itself, which is determined directly from the screen voltage and screen current, and in addition there is a possible loss depending upon how the screen is fed. If, for convenience, we feed g_2 from the anode supply through a dropping resistance as shown in Fig. 177, we shall have an I^2R loss in this resistance. If the screen is supplied from a separate source we avoid this loss, but this method is not too satisfactory, because when the valve is modulated, the screen current will vary (as has been explained), and this produces distortion.

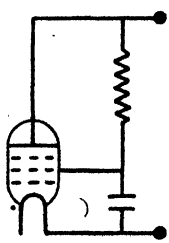


FIGURE 177.

In order to obtain full emission for anode and screen the cathode has to be rather larger than for a triode of corresponding size.

Considering the problem in general terms, therefore, it will be found that the overall efficiency of a pentode transmitting valve is not materially different from the triode counterpart as the gain of grid driving power is offset by screen and cathode losses, and the following table shows comparative figures for a triode, pentode, and screen-grid valve of similar size as regards power output and voltage of supply.

TABLE II

Valve.	Angle of Current Flow.	Drive Power.	Screen Loss.	Power Input.	Total H.F. Power.	Conversion Effy.	K.W. to Load.	Overall Effy.	Stage Power Gain.	Volts.
Triode	156°	·035	—	1·43	1·03	72%	1·0	69%	27	4000
Tetrode	150°	·025	0·27	·97	·68	70%	·6	66%	27	4000
Pentode	150°	·020	·114	1·90	1·43	75%	1·27	69%	63	4000

Pentode Power-Amplifier Stage. A normal stage of power amplification using a pentode valve is shown in Fig. 178, where L_gC_g is the tuned input (driven from a previous stage) and LC is the tuned output-circuit. For the time being it should be assumed that the suppressor grid is at zero potential, the condition which gives nearly maximum output, and that

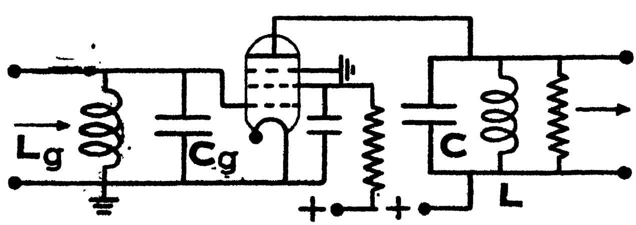


FIGURE 178.

the screen potential is supplied from a separate source through a small series resistance. It will be observed that no neutralising circuit should be necessary provided the mechanical construction of the system is correctly carried out, so as to remove

coupling between grid and anode circuits. This involves the screening of the resonant circuits one from the other and the correct mounting of the pentode valve within a screen.

It will be observed that since there are no neutralising arrangements provided, there is no adjustment by which instability of the pentode transmitter can be prevented as there is with a triode stage. This does not mean, however, that the pentode circuit is inherently stable. In fact, many precautions need to be taken when designing pentode circuits, and checks must be made to determine that the circuit set

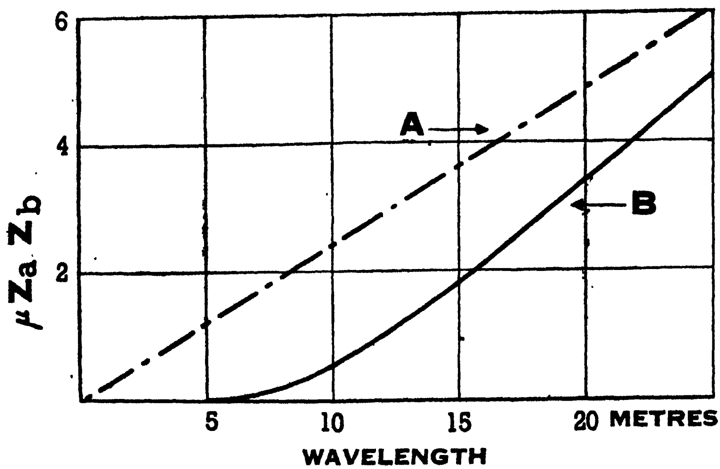


FIGURE 179.

up is free from incipient oscillation at the driving frequency and also from parasitic oscillations.

Considering the stage such as shown in Fig. 178, it is found that if the impedances in the anode and grid circuits are too high the stage becomes unstable and the lower the wavelengths the greater the instability, the connection between wavelength and stability being as shown in Fig. 179. Curve "A" is a theoretical curve which assumes that the "carry-through" capacity is mostly C_4 and that there is no inductance in the g_2 and g_3 leads. If the value of $\mu Z_a Z_b$, at a given wavelength, is greater than that shown on the curve, then the simplified theory states that the amplifier circuit will self-oscillate. Curve "B,"

the shape of which will vary with the type of valve and circuit layout, is typical for a small pentode valve suitable for about 100 watts output. The divergence in the two curves is due to the fact that there is "carry-through" via the other electrodes which cannot be truly shorted to earth potential. Thus considering Fig. 180, which shows the effective circuit, it will be seen that although the screen and suppressor grids are ostensibly earthed directly through their shunt condensers (not shown) the effective inductance of their leads will have the effect of producing a carry-through voltage between the electrode and earth, and in consequence from anode back to the control-grid.

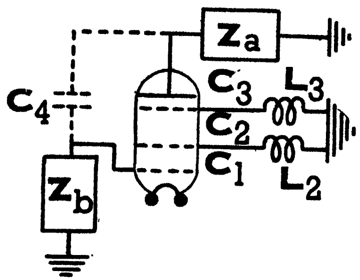


FIGURE 180.

For the particular valve in question, the relative values of $\mu Z_a Z_b$ are given on the curve, and the values of $L_2 L_3$ and the inter-electrode capacitances are as under :

$$L_2L_3 = 4 \times 10^{-15}\ H.$$
$$C_1 = 25\ \mu\mu\text{F}.$$
$$C_2 = 30\ \ ,,$$
$$C_3 = 50\ \ ,,$$
$$C_4 = 0{\cdot}05\ ,,$$

Although the absence of neutralising simplifies the circuit and reduces the possible number of parasitic paths, such oscillation can occur unless precautions are taken in the layout of components and all leads kept as short as possible. The various modes possible which are all of the ultra-short wave type can be realised by studying Fig. 181, where the heavy lines indicate possible circuits, it being assumed that the leads shown all represent small inductances.

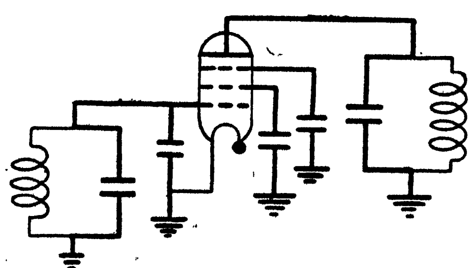


FIGURE 181.

As the wavelength is reduced, the tendency to self-oscillate at an ultra-short wavelength increases, and influences the general shape of the wavelength-output characteristic somewhat as shown in Fig. 182. The rise of output at "B" is due to a phase of reaction which assists the drive, whereas the sharp fall at "A" is due to a negative feedback condition arising.

It was mentioned previously that with a pentode stage there is no preliminary check of the stability of the stage, nor of the correct tuning position of the output stage as there is

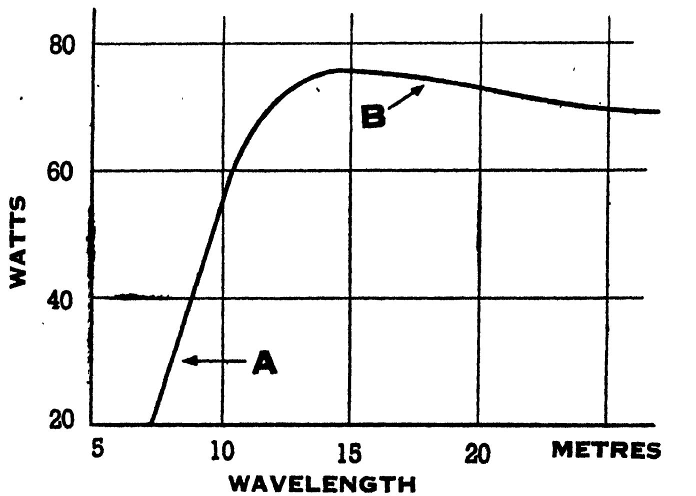


FIGURE 182.

with a triode amplifier (see page 285). When the anode supply is on, the change of anode and total feed currents and grid current is much less than with the triode, as can be seen by comparing the curves of Fig. 183 with the corresponding curves for the triode shown in Fig. 170.

The Design of the Output Circuit. In transmitter design it is sometimes a matter of difficulty to decide whether the output circuit, coupling from the transmitter tank-circuit to the feeder or aerial should be of the series or parallel form, shown schematically in Fig. 184.

When the load resistance is low, say up to 100 ohms, the series type of circuit will always be adopted as it is easy to

obtain full loading from the transmitter at unity power factor. If the load is of high resistance, say above 500 ohms, the parallel circuit will normally be found easy to design, as in this case the load resistance is high compared with the reactance of the coupling coil (and of the parallel tuning-capacitance), and full loading can be obtained at a unity power-factor.

For intermediate values of load resistance it is more difficult to decide which type of circuit to use. If the load is connected in series, then, as it increases in resistance a larger induced E.M.F. is necessary to circulate a given power. If the coupling coil has already been arranged so that the E.M.F.

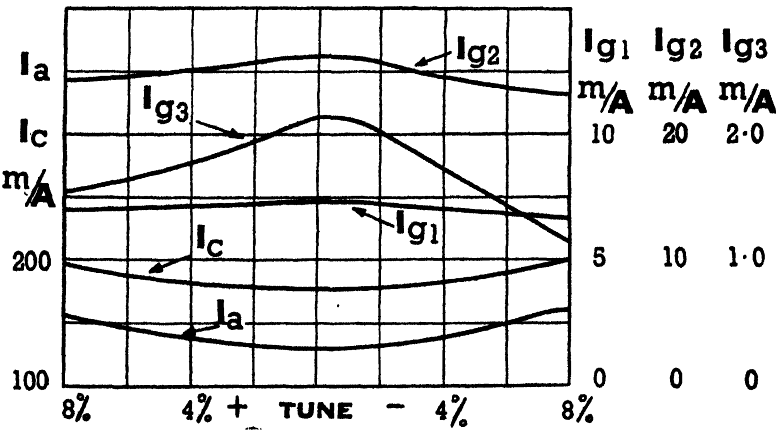


FIGURE 183.

induced in each turn of it is as great as possible, it will be necessary to increase the turns. At the higher frequencies this may result in the coil resonating on its own, due to its self-capacity.

If the parallel connection of load is employed, it can be shown that unless the coil reactance is less than half the value of the load resistance, the value of capacitance which makes the current in the load a maximum will not produce resonance —that is, the total reactance in the circuit will not be zero. For values of load resistance between about 100 and 500 ohms, therefore, it will be necessary to consider each case on its own merits and decide whether the series or parallel connection will be the most economical.

General Design Data. The design of short and ultra-short wave transmitting apparatus calls for special attention, both to the materials used and the design and layout of the components. All circuits must be carefully screened from one another by good quality copper, brass, or aluminium sheets, carefully bonded to each other and earthed to a definite earth busbar. If this is not done not only will there be interaction between the circuits, but the proximity of any H.F. circuit to surrounding dielectrics, such as wood or plaster walls, will increase the losses very considerably, even if no dangerous heating is set up in the dielectrics themselves. The H.F. circuits must be designed so that as few insulating

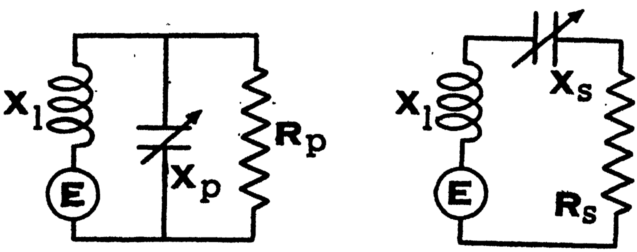


FIGURE 184.

materials are used as possible, and where they must be incorporated they should be placed as much as possible outside the field. Connecting leads must be kept as short as possible and of ample dimensions. Apart from the general screening of the H.F. apparatus, supply leads to filaments, grids, etc., should be run in screened compartments or tubing as near to their final connecting point as possible. No end or short-circuited turns should be used when wave-changing is required, but either separate coils designed for each wave, or an even number of single turn coils used, which can be connected in series or series parallel, or all-parallel, to give a variety of inductance values. This method is very suitable, as the conductor size automatically increases as the wavelength decreases.

With high powers and voltages it is necessary to avoid all sharp corners and edges to prevent brushing, or what is known as a "torch" discharge. At very high frequencies the energy

supplied to the air by a brush discharge is much greater than at lower frequencies, general ionisation of the air around takes place and a flash over occurs which is more in the nature of an arc at such frequencies. The arc may flash over to another electrode as with such discharges at lower frequencies, but it will often be found to arc out into space, when it is known as a torch discharge. To avoid these discharges, all conductors carrying H.F. current should be designed with rounded edges, and this means, with condenser plates, either making them with a heavy section, or built up with a hollow section, an example of the latter type of construction being shown on page 517.

Copper, tubular coils are suitable, or flat copper strip with the flat surface parallel to the axis of the coil. A very convenient design is the square-shaped inductance used by Franklin (see Fig. 313), as this lends itself not only to mechanical design, but to easy cross-connecting and fine adjustment of individual turns, without giving any detrimental end-effect.

The spacing of turns is usually not greater than twice the diameter (or width) of the turns, and very little alteration of inductance is obtained by altering the spacing, because the end-to-end capacity is of greater consequence than the increase of mutual inductance between the coils.

Dielectrics for High Frequencies. A considerable number of insulating materials suitable for use at high frequencies have been produced in recent years as many of the materials which are efficient at low frequencies have far too high a dielectric loss at high frequencies, since this tends to be a constant per cycle of applied voltage.

The effects of a high dielectric loss are, of course, that the effective resistance of the circuit concerned is increased and the dielectric may become distorted if not destroyed. In the case of materials to be used as insulating supports, the main requirements are mechanical strength, low dielectric loss and high dielectric strength, the dielectric constant being relatively unimportant. They should be non-hygroscopic. Amongst suitable materials is "Mycalex" (made from mica and glass), as this can be obtained in a variety of forms, and lends itself to drilling, milling, and ordinary machine processes. Other materials are "Frequentite" and "Calan," which are manufactured in the same way as other ceramic materials, the former

being made from steatite and the latter from finely divided mica. Of the numerous synthetic resin insulators "Trolitul" has excellent properties, and like other plastics can be moulded into complicated forms. Porcelain, although it is very uneven in character, is still extensively used, particularly in places outside the H.F. field, but only those types which contain no filling material except lead are really suitable.

When dielectrics are wanted for building up condensers, a low dielectric loss is evidently desirable, mechanical strength is not so necessary, but a high dielectric constant is useful in order to reduce the size of the condenser, this being especially useful because the small size reduces the inductance of the leads and plates themselves.

In short and ultra-short wave transmitters, the tuning capacity is almost always a low-capacity, variable air-condenser, but for blocking condensers mica sheets of good quality are still much used. Recent Continental research has produced some remarkable materials having enormous dielectric constants. These are being increasingly used in receiving components and will no doubt become more extensively used in transmitting equipment. The properties of some dielectrics are tabulated below, and the table shows the notable advance of these materials over the rubber compounds such as ebonite, the standard insulating material of a few years ago.

TABLE III

Material.	Dielectric Constant.	Power Factor.	Dielectric Strength kv/mm.
Calan	6·5	·0004 at 10^7	40
Frequentite	6·0	·0008 at 10^7	50
Mycalex	6·0	·003 at 10^6	14
Mica	7·0	·0002 —	50
Trolitul	2·5	·0003 at 10^6	30
Porcelain	5·5	·008 at 10^6	—
Fused quartz	3·8	·0002 —	20
Ebonite	3·0	·009 at 10^6	150
Kerafar	80·0	·001 at 10^6	—

Extracted from the "Journal of Scientific Instruments," Vol. 15, page 217. Values obtained by Dr. Hartshorn at the N.P.L.

Valves for Very High Frequencies. It is clear that the successful development of short and ultra-short wave apparatus depends largely upon the valve design, which has had to go hand-in-hand with the circuit technique. As we try to build power amplifiers for increasingly high frequencies, a number of difficulties arise, and the successful exploitation of ultra-high frequencies largely depends upon these difficulties being overcome if conventional types of circuit known to be satisfactory at the lower frequencies are to be used.

Evidently a compact layout and the avoidance of stray coupling and long connecting leads helps to produce circuits which will tune to a very high frequency, but all the care that is taken over the circuit can be upset by the valve design. For ultra-high frequency work, valves should have short, direct connections between the electrodes, and the seals through the envelope should be massive, not only to carry the high frequency currents but to reduce the lead inductance to a minimum. Inter-electrode capacity must be kept as small as possible, although the spacing between the electrodes must be kept near together in order to reduce the transit path of the electrons across the valve.

Normally we assume the transit time of electrons through the valve to be quite negligible and suppose that the electron stream follows instantaneously a change of grid voltage. At a very high frequency, the value of which depends upon the size of the valve, this assumption is untenable since the transit time of electrons from cathode to grid and grid to anode becomes an appreciable fraction of a cycle. One effect of this will be to increase the effective conductance between grid and cathode, thereby increasing the load placed upon the driving circuit. A second result is that although the maximum electrons leave the cathode when the grid is at its maximum positive value, yet they do not arrive at the anode when the volts there are a minimum (examine Fig. 162, where transit time is assumed to be zero). As a result the losses are increased, although the effect can be partly compensated for by detuning the anode load-circuit slightly.

The above difficulties become more serious when we try to increase the power rating of our amplifier. The size of circuit components has to increase in order to carry the currents,

spacing must be greater to stand the greater voltages, and the valve dimensions have to be increased, thus increasing the transit time. For instance, in a normal design of water-cooled valve designed for short wave working, the transit time from cathode to grid is $\cdot 65 \times 10^{-9}$ seconds, that from grid to anode $1 \cdot 3 \times 10^{-9}$ seconds, or the total time from cathode to anode 2×10^{-9} seconds. At an ultra-high frequency of 5×10^{7} (i.e. 6 metres) this represents one-tenth of a cycle of oscillation.

SELECTED REFERENCES.

(1) *Vacuum Tubes as Power Oscillators.* Prince. *P.I.R.E.*, Vol. 11; 1923.

(2) *Simplified Methods for Computing Performance of Transmitting Tubes.* Wagener. *P.I.R.E.*, Vol. 25; 1937.

CHAPTER XI

OSCILLATORS AND CONSTANT FREQUENCY OSCILLATORS

ALTHOUGH the discovery of the utility of short waves greatly extended the useful "wireless spectrum," at the same time it so greatly increased the usefulness of wireless communication that the new frequency bands were quickly filled up. Since the modulation band width required is no greater with short than with long waves, it follows that if similar selective methods of reception are to be used and the available band of frequencies fully exploited, the permissible frequency variation (expressed as a percentage) becomes much smaller. The frequency of the ideal transmitter would conform to four quite distinct requirements:

(1) It would be correct.

(2) It would not "drift," that is, change slowly with time.

(3) It would not "scintillate," that is, change from instant to instant.

(4) It would be possible to alter it if occasion arises.

In order to obtain the first condition, accurate methods of absolute frequency measurement are evidently necessary, and these have now been so highly developed that measurements of frequencies can be made to less than one part in a million.

With regard to the second condition, several different types of constant frequency sources are now available and will be discussed in this chapter. Such circuits produce only a very small output and a chain of amplifiers is required before sufficient power is available to drive even a small transmitter. It has already been pointed out (Chapter X) that a self-oscillator is but seldom used as a transmitter for short waves owing to its general frequency instability, and the main features of a driven system were discussed in Chapter X.

The third requirement is bound up with the design of the complete transmitter. Although a source of frequency having a very small drift may be provided, it by no means follows that the frequency radiated from the transmitter will be perfectly steady. It may be found to vary at a rapid rate as soon as the transmitter is modulated or keyed. This is because the changing loads on the transmitter are reacting back on to the frequency source in some way. To prevent this the driving source should be kept oscillating and drive into a small but constant load at all times. This is best accomplished by arranging the first amplifying stage, the isolator, to be unaffected by subsequent modulated stages, and if possible operating the valve of this amplifier without grid current.

Further, if the master oscillator be run at a lower frequency than the main power amplifier, this in itself will tend to prevent reaction effects. There appears to be an idea prevalent that it is desirable to produce master oscillator circuits at the same frequency as the final frequency to be radiated. Actually it is a considerable advantage to employ one at a lower frequency, not only because of immunity from reaction effects, but because a larger choice of final frequencies becomes possible from one master oscillator. The fourth requirement is a difficult one to cater for, but in some cases a necessary one to provide and may prove the deciding factor in choice of master oscillator to be adopted.

It is now possible commercially to produce transmitters the frequencies of which do not drift by more than 1 part in 10^6 over very long intervals (although a tolerance of 1 part in 20,000 is sufficient for most purposes) and which scintillate by an amount too small to measure.

In the case of short wave telephony it is essential to reduce the scintillation to very small proportions, apart altogether from the question of interference with other channels, otherwise bad distortion due to multiple echoes is produced. We have, therefore, to consider suitable types of circuit to generate a very constant frequency and capable of delivering sufficient output to drive a chain of amplifiers. Before the need for such great precision arose, earlier driven systems were controlled by a fairly large (100 watt) drive, consisting of an ordinary valve self-oscillator, working at the frequency to be

radiated, but it was found difficult to design a stage of this size to be free from frequency drift, particularly on first starting up.

As the permissible tolerance became less, wireless engineers looked around for means of obtaining a more stable driving source, and three distinct types of "constant frequency" drives have been evolved, all of which are maintained in oscillation by valves, but the resonating element may be:

(1) Tuning fork or steel rod.
(2) Piezo-electric crystal.
(3) Special electrical resonant circuit.

The valve maintained tuning fork, due to Eccles, was the first type developed for long waves, and since this is of necessity a low-frequency source very many frequency multiplying stages are necessary before it can be of use to control a short wave circuit.

The second type, developed by Cady and Pierce as a result of original work by the Curie brothers, makes use of the piezo-electric properties of quartz, tourmaline, or Rochelle salt. Such crystals can operate efficiently up to 15 megacycles, and thus but few frequency multiplying stages are necessary even for short and ultra short-wave circuits.

The use of mechanical oscillators temporarily eclipsed the resonant-circuit oscillator largely on account of the effectively high Q value which is obtained by such means and the ease with which a circuit of zero temperature-coefficient can be obtained, but recent intensive work on the technique of the electrical resonant circuit oscillator has established it again in the fore-front, and for certain purposes it compares favourably with mechanical drive systems.

The choice of what type of master oscillator to use is often difficult. All types of constant-frequency, master-oscillators are expensive, the main item not being that of the actual components used but the cost of final setting up and calibration, this rising rapidly as the tolerance is reduced. On short waves the fork is but little used, and the choice lies between the crystal and an electrical resonant circuit. Where very great precision is required, and the frequency can be decided without doubt, the crystal will usually be chosen, but if there

is any question of flexibility, then the electrical resonant circuit will be preferred.

Quartz. Certain crystalline substances, notably quartz, tourmaline and rochelle salt, exhibit a phenomenon known as the piezo-electric (or pressure-electric) effect. Of these, rochelle salt, although the most active, is unstable physically and it is not greatly employed. Tourmaline is used occasionally,

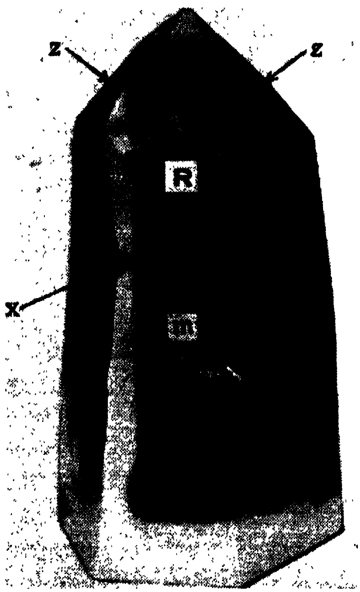


FIGURE 185.

but for most wireless purposes quartz is preferred and we shall only consider this. Quartz is a very common form of silica, but good crystals are only found in quantity in Brazil, Madagascar, and Japan, and as only a very small percentage of the quartz found is of any value for piezo-electric work, very great care is required in the selection of the raw material if large wastage is to be avoided.

A natural quartz crystal can appear in various forms but an idealised crystal should take the form of a hexagonal prism terminating at both ends in a pyramid. The quartz found,

however, is far from perfect and in many cases badly deformed with faces of uneven size and one end larger than the other. Generally speaking, specimens are found broken and a natural crystal will usually taper as shown in Fig. 185.

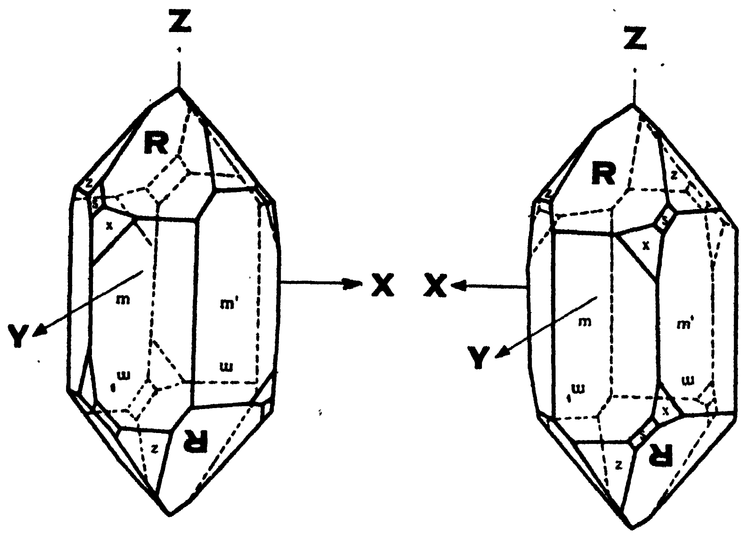
Nomenclature. Much confusion arises in dealing with the nomenclature of quartz, as not only have crystallographers not been consistent, but various investigators in piezo-electric work have each produced their own particular system of notation which appeared to be best suited to their needs. This means that the reading of current literature is difficult. In an attempt to get rid of the confusion arising, a notation was suggested by the National Physical Laboratory,[2] but it has not been widely adopted, possibly because its manner of presentation is more suited to the scientist than to the practical works engineer. In the sections which follow we propose to adhere largely to the suggested N.P.L. notation but to correlate it, where it appears desirable to do so, with notation adopted by other workers.

Types of Crystal. Referring to Fig. 186, which shows idealised types of double-ended crystals, the main body contains three pairs of parallel faces designated m, m', capped at each end by a hexagonal pyramid, the six faces of which are denoted alternately by the letters R and z. The position of the crowns at the opposite ends of the crystal is such that the R (and z) faces are above one another, but the crown is skewed so that the R face of one is directly above the z face of the other crown. We may really regard the R and z faces of the crowns without the body as forming the sides of two interleaved rhombohedrons. Although the size of the various faces varies considerably, their angular relationship is very definite, however distorted the crystals may be, and this angular relationship can act as a guide when making cuts. Thus the included angle between adjacent m m faces on the basal plane is 120°, and the angle between an R face and the optical axis is 41° 47′.

In certain crystals, facets s and x occur at the junction of the pyramid as shown in Figs. 185 and 186, and the order of these facets gives the clue to the form of crystal. The surface of a crystal usually shows growth lines across the mm' faces, seen in Fig. 185, which in a good crystal are parallel and unbroken.

Quartz may assume one of two simple forms which are the mirror image of each other as seen in Fig. 186; or a complex crystalline structure may result, known as twinning, the latter type of crystal being of no value for piezo-electric work.

Examples of perfect twin crystals are shown in Fig. 187, but it should be explained that internal twinning may be present in a crystal which to outward appearance is of simple form.



Left-handed crystal.

Right-handed crystal.

FIGURE 186.

The notation for crystal faces is shown in Table 1 and the position of these faces can be seen by reference to Fig. 186.

TABLE I.

N.P.L.	R.C.A.
R	A
z	B
s	C
x	D
m (under *R* face)	E
m (under *z* face)	F

NOTE.—For convenience in discussion we propose to denote the *m* face under the *z* face as *m′*.

The simple forms, so called right and left-handed, both of which are equally suitable for piezo-electric work, are determined by whether the sequence of facets *m*, *x*, *s*, and *z* follow the course of a right-hand or left-hand screwthread. Or more simply, by whether the facets *s* and *x* are to the right or left of an *R* and *m* face, when viewing the crystal with the *R* and *m* faces to the observer. Such an external feature indicates a molecular structure which will give a right-hand or left-hand rotation to the plane of polarised light transmitted along the optical or *Z* axis.

Thus if plane-polarised monochromatic light is passed through a slab of right-handed crystal cut normal to the optic

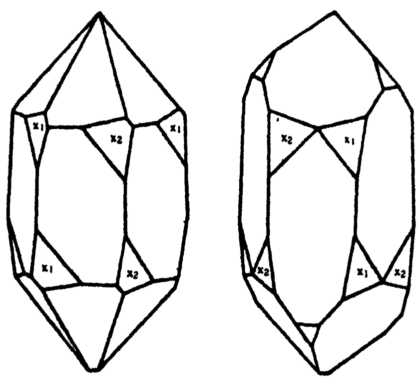


FIGURE 187.

axis, the plane of polarisation is rotated in a clockwise direction when the observer looks at the source through the slab. It is also the direction in which a microscope analyser must be rotated by an observer who to restore light looks through the analyser towards the source. Amongst certain American workers a reversed convention is used but it will become evident later that, provided due precautions are taken, the "hand" of the crystal is quite immaterial.

The principal axes of the crystal are known as the *Z*, *Y*, and *X* axes, which are all at right angles to each other. The *Z* or optical axis, which has already been mentioned, is that about which the rotary polarisation of light takes place. This axis is the same as the crystallographic axis of three-fold symmetry, because the structure repeats itself three times around it,

and a knowledge of the axis is useful in the selection of suitable specimens and subsequent cutting.

The *X* or electrical axes, of which there are three, are each one parallel to an *m* face, and they have a two-fold symmetry.

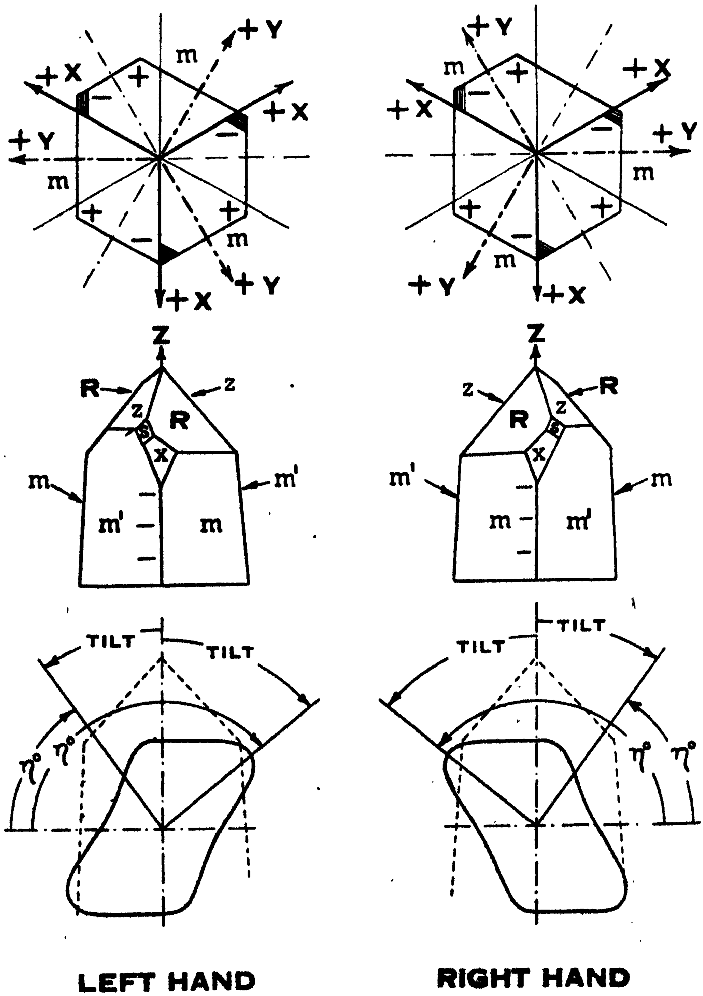


FIGURE 188.

Thus in Figs. 188 and 189, the thick line axes, marked *X*, each at 120° to each other, are electrical axes having a positive sense. It is to be noted that only in a perfectly equal-sided

hexagonal crystal will the X axes pass through the corners of the crystal.

At right angles to the electrical axes are the so-called mechanical axes, of which again there are three (shown chain dotted and marked Y), each one of which emerges from an m face with positive sense, and is at right angles to an X axis.

With the N.P.L. notation, the rotational sense of a positive X axis to a positive Y axis is the same for both a left-hand and a right-hand crystal as shown in Fig. 189, but the usual mathematical notation is such that the rotational sense of the

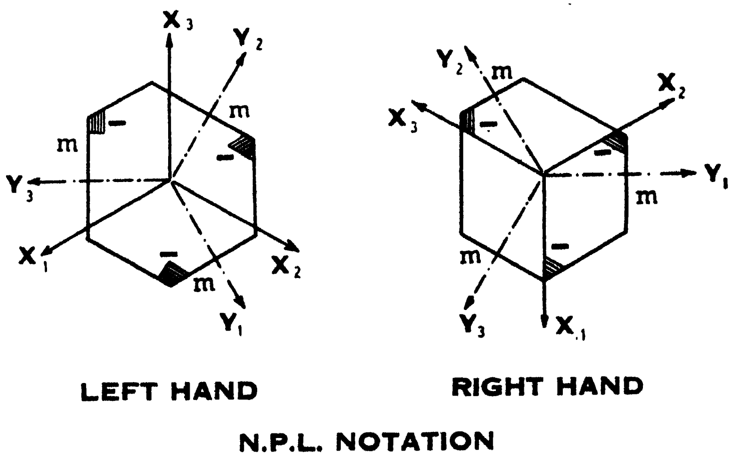


FIGURE 189.

positive Y axis to the positive X, is clockwise for a right-handed, and anti-clockwise for a left-handed crystal, looked at from above, as shown in Fig. 188.

Actually, the point is not very important because, when orientating a crystal about the Z axis, the same results are obtained if it is orientated from a given axis (i.e. a point of symmetry), either clockwise, or anti-clockwise.

The Piezo-Electric Effect. When stressed by pressure or torsion, a quartz crystal may, in certain circumstances, exhibit a piezo-electric effect, that is a mechanical force may produce an electric charge. Or conversely, when placed in an electric field it will exhibit strain and mechanical distortion.

Thus if a quartz crystal is compressed by pressure along one

of its *X* axes, which implies an extension along a *Y* axis at right angles, the six edges of the crystal become charged alternately positive and negative, a negative charge always appearing at those edges carrying facets, or in the absence of facets, those edges which would have carried them had the crystal been so formed. The edges in which the compressed axis terminates exhibit the greatest charges; thus in Fig. 188 (and 189), if pressure is applied across a pair of edges, say front to back (elevation) negative charges will appear on the three edges under the facets, shown shaded in plan, both with the left and right-handed crystal. Such charges of course could be observed by means of a low-capacity electrometer.

It will be observed that when the position of the *R* and *m* face is determinable, even if no facets are present, it is possible to find the hand of the crystal by applying pressure along any *X* axis and noting the sense of the induced charges at the corners. As is seen from Fig. 188, if the induced negative charge is at the right-hand edge of an *m* face, the crystal is right-handed.

It is clear that the application of pressure along a *Y* face will have a similar effect (although it is not a usual procedure), except that the charges induced will be opposite in sign, but pressure applied in a direction along the *Z* axis produces no electric charge anywhere, neither does pressure applied evenly to all parts of the crystal (such as hydrostatic pressure), and the application of pressure to the crystal in directions other than *X* and *Y* will produce, as would be expected, complex, distributed charges.

As stated above, the converse effect also occurs, that is, the application of an electrical potential across an axis in the equatorial plane produces a contraction or extension along corresponding *X* and *Y* axes. Hence if we imagine a plate of quartz cut at an appropriate angle, subjected to an alternating electric field, it would be forced to expand and contract at the frequency applied. The mechanical oscillation will be of very small amplitude, however, unless the applied frequency is the same as the mechanical resonance of the plate which depends upon dimensions, the way in which it is cut, and upon the elastic constant of the quartz. We will see later that the frequencies of oscillation that can be produced are extremely

high and thus we have a means of producing a convenient form of mechanical oscillator which can be maintained electrically at radio frequencies and piezo-electric control devices are becoming of increasing importance in wireless work. Before discussing the variety of ways in which crystal plates may be cut, it will be desirable to say something about the selection of suitable material.

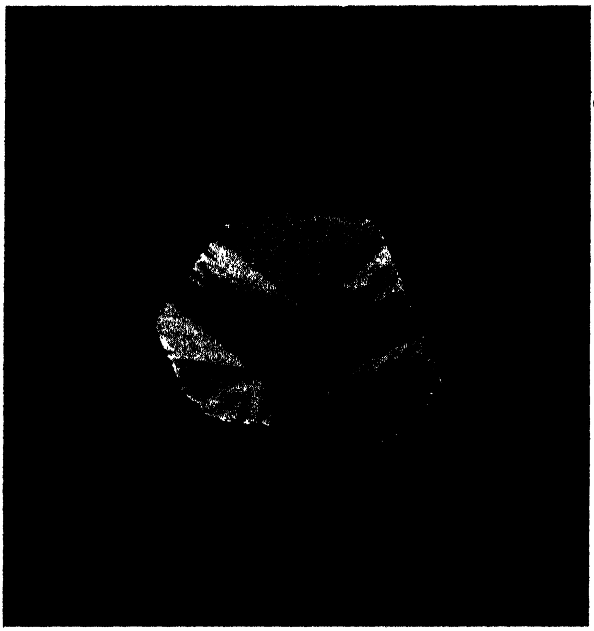
Selection of Quartz. Since a large percentage of the natural quartz is useless for piezo-electric work, methods of preliminary inspection and test are of considerable importance. Although in some cases the external shape can give a clue to the goodness or otherwise of the crystal, it is often difficult even to distinguish the faces one from another on account of malformation. To give an example: out of 10 selected samples, with a possibility of 30 x and 30 s facets, only 5 x and 10 s were actually present. Further, quartz which to the eye appears a simple growth, as Fig. 185, may internally be malformed. Hence optical and electrical tests become desirable before any cutting takes place.

A preliminary examination of the raw quartz is often made by placing it in a bath of winter green (Methyl-salicitate). Plane-polarised monochromatic light is projected along the Z axis, and the crystal viewed through an analyser along the Z axis, which will detect twinning. A further and standard test for the detection of one form of twinning is to polish and etch with hydrofluoric acid selected faces or sections, and examine the etched surface pattern by eye.[11] Having eliminated crystals with twin growth, simple types having mechanical flaws are best avoided.

Having selected a specimen which is probably of good quality, the crystal will be sliced through normal to the Z axis and a specimen slice taken for optical examination using a polariscope.

A polariscope consists of a source of light, polarised by a Nicol prism, in line with which is a second Nicol, and a viewing eyepiece. The second Nicol is rotated so that light is extinguished, and if now the quartz section be placed between the Nicols, because of the optical properties of the quartz, light again passes in a manner which reveals the structure of the slab.

Suppose a perfect specimen of crystal, which has been cut exactly normal to the Z axis, to be placed in a polariscope in which a "white" light is used. Then the crystal would appear to be of uniform colour when viewed through the polariscope, the actual colour depending upon the thickness of the slab. When an average specimen is viewed, however, it will usually be found that only an area near the centre will



QUARTZ SLICE UNDER POLARISED LIGHT.

FIGURE 190.

show uniformity and the rest of the section will be broken up by a spectrum of colours whose brilliance and shape reveals the presence of imperfections in the main crystal. Fig. 190 shows a slice photographed through a polariscope, where the light parts represent the faults and are seen as a series of colour spectra. These areas are quite useless and will be marked so that they can be ignored in the subsequent cutting process. For the cutting of inclined plates (to be discussed later), it is essential to be able to identify the original positions

of *R* and *z* faces with respect to the *m* faces. If this was not done by studying the pyramid top, before cutting the specimen slice (possibly because of malformation), the various faces can be found by applying the electrical test mentioned on page 314, by viewing the slab through a polariscope, and by determining the hand of the crystal. This can be done by the use of plane-polarised white light viewed through an analyser and the employment of a single or double bi-quartz wedge. The electrical test determines the corners at which the *s* and *x* facets were above, and the hand of the crystal as determined by the polariscope and bi-quartz wedge determines whether the *m* face is to the right or left of the negatively charged corner as shown in Fig. 188. Thus although a knowledge of the hand of a crystal is of no importance in itself, it becomes necessary as a means of distinguishing between the *m* face below an *R* face and an *m* face below a *z* face.

As a result of the optical and electrical tests which have been indicated, it is possible, not only to pick out the areas of bad quality from a crystal specimen, but also to identify the faces, and so obtain a datum from which to cut a crystal to a specified angle, no matter how malformed the original specimen may have been.

Cutting. It can be imagined that with crystals of uneven shapes and weighing up to several pounds, the holding of a crystal so as to obtain a cut of specified angle to any given axis or face is a difficult matter, and up to now it is the usual practice not to attempt direct cutting, but to saw the block up into slabs cut normal to the *Z* axis and cut from such slabs as required. Cutting is carried out on what is virtually a milling machine, using for a cutter a type of lapidary's saw, that is a thin copper disc charged at the rim with diamond dust; or thin mild-steel discs are also used; or thin bakelite discs impregnated with carborundum or diamond. For sub-dividing into the finished crystal it is the practice to mount the rough blocks on to an accurately ground face-plate (often with thin ground-glass supports above and below) by an adhesive gum, and cut from the block the required crystal plates. A sufficient margin of material must be allowed for lapping to the finished size, the margin necessary being determined by the type of crystal-cut and the final precision required. The object of

the glass supports is to prevent chipping of the crystal edges.

Although the final lapping of high precision crystals is always carried out by hand, there has been a great advance made in the use of automatic lapping machinery, following the technique used in producing high-grade lenses. A great deal depends upon the care and adjustment of the machines themselves, but it is possible to obtain with a good machine a surface which is flat to within a tolerance of 10^{-4} of an inch. When a fine polish and true surface is essential, a final lapping is carried out by hand using a stationary lap, and it will be

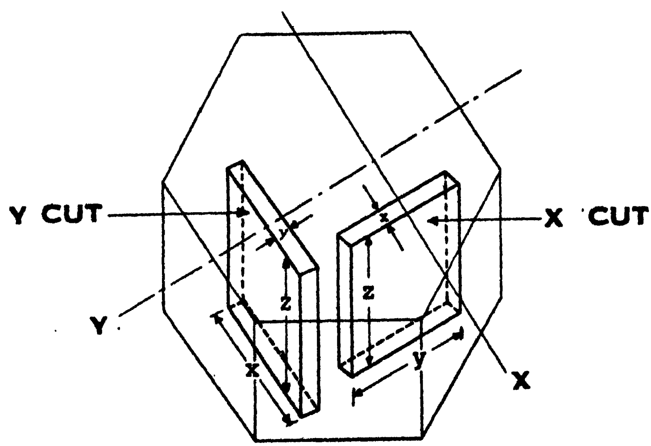


FIGURE 191.

appreciated that in high-precision work it is impossible to work by micrometer readings alone, and it becomes a matter of frequent rubbing, cleaning, and testing in situ against a precision frequency-standard.

Types of Cut of Crystal Slices. We propose to discuss first, plates of the simple so-called Z, X, and Y cuts, whose main definition is generally agreed by all writers, and to which all other types of cut are directly related.

Consider a thin, rectangular crystal-plate cut from a slab such as we have been discussing. When the cutting plane is normal to the Z axis, the resulting cut is called a Z cut, and a plate so produced will have no piezo-electric properties if the electrodes also are normal to the Z axis, as has been explained.

When the cutting plane is made normal to an X, or electrical axis, the main faces lie in the YZ plane, and a parallelapiped so cut would have sides of lengths x, y, and z, where these are dimensions parallel to the axes X, Y, and Z, as shown in Fig. 191 (right). Such a crystal plate is known as an X cut; or sometimes as a "Curie," or "face perpendicular" cut, terms now to be deprecated. Piezo-electrically such a crystal will have a strong edge vibration along the Y axis, and a weaker mode of compression and extension along the X axis. Since the face dimensions are usually large compared with the thickness, the former frequency will be low compared with the latter, the rule connecting frequency with crystal dimensions being approximately:

$$f_{\text{edge}} = \frac{2750}{y} \quad . \quad . \quad . \quad . \qquad (1)$$

$$f_{\text{thickness}} = \frac{2750}{x} \quad . \quad . \quad . \qquad (2)$$

where f is in kc/s and y and x in mm.

If we cut a similar thin rectangular plate lying in a plane normal to a Y axis, Fig. 191 left, such a plate is known as a Y cut, and sometimes as a 30°, or face parallel cut, terms now to be deprecated. Such a crystal will be found to be very active in the Y direction, that is in shear mode about Z, the resulting (thickness) frequency being:

$$f_t = \frac{2070}{y} \quad . \quad . \quad . \qquad (3)$$

where f is in kc/s and y is in mm.

The reason for the 30° classification can be seen by reference to Fig. 188. From this diagram it is clear that rotation of a vertical cutting plane about Z alternates every 30° between an X and Y plane, and hence from an origin of an X cut, the adjacent Y cut will be obtained by rotation of the cutting plane by 30°.

The N.P.L. notation reclassifies X and Y cut plates, as the vertical cutting plane is considered to have an origin along the X axis (giving a y cut) and thus a 30° rotation from this origin results in an X cut crystal plate.

It should be noted in passing that certain workers consider adjacent X and Y axes as complementary, whereas others

prefer to consider the complementary X and Y axes at right angles.

The formulæ 1, 2, and 3 are derived from a consideration of the mechanical structure of the crystal. The oscillation of a quartz plate is due to mechanical waves in the plate setting up stationary waves, either longitudinally or transversely to the mechanical axis.

The velocity of propagation for a mechanical wave is given by

$$v = \sqrt{\frac{E}{\rho}} \quad . \quad . \quad . \quad (4)$$

Where v is velocity in cms. per second,

E is the elastic modulus in dynes per cm^2 in the plane considered.

ρ is density in gm. per c.c.

For normal specimens of quartz

$$\left.\begin{array}{ll} & E = 8 \times 10^{11} \\ \text{and} & \rho = 2{\cdot}654 \end{array}\right\} \text{Average values.}$$

Then $v = 5{\cdot}5 \times 10^5$ cms. per second.

Now $\lambda = \dfrac{v}{f} = \dfrac{5{\cdot}5 \times 10^5}{f}$ where λ is in cms. and f is in cycles per second.

or $\lambda = \dfrac{5500}{f}$ where λ is in mms. and f is kcs per second.

If the crystal is free, the fundamental vibration, whether longitudinal or transverse, will clearly be at a half wavelength with a node of movement at the crystal centre. Thus if y and x are crystal dimensions in mms. along the Y and X axes, then since:

$$\lambda = 2y \text{ and } \lambda_1 = 2x$$

$$2y \text{ or } 2x = \frac{5500}{f}$$

$$y \text{ or } x = \frac{2750}{f} \quad \text{as stated above.}$$

This example applies to an X cut crystal only.

Neither the X cut, nor the Y cut, crystal plates are now greatly used as both have a poor temperature-coefficient, and suffer from a phenomenon known as stepping, to be explained,

due to the complex type of vibration within the crystal. With thin plates, where the thickness is not more than $\frac{1}{50}$th the edge length, the temperature-coefficient of the X cut plate is of the order of -20 to -50 parts in 10^6 for $+1°$ C. temperature rise, and that of the Y cut, of the order of $+60$ to $+100$ parts in 10^6 for $+1°$ C. temperature rise, the thinner the plate cut, the worse the value of temperature-coefficient.

Although the thickness frequency of a Y cut plate varies approximately as $f_{kc/s} = 2000/y$, yet if a plate be gradually decreased in thickness, a curve of f against y does not give a straight line but sudden jumps in frequency occur at points to give a discontinuous line. These discontinuities are termed stepping points, and result when the edge vibrations, controlled by the size of the faces of the plate, are coinciding with sub-multiple frequencies of the wanted thickness frequency.

We can really liken a quartz plate to a very high Q electrical resonant circuit (values of 25,000 in air and 200,000 in vacuo being possible), to which is coupled a number of subsidiary tuned circuits, also of high Q; the degree of coupling and frequency relationship to the fundamental varying with crystal dimensions. As long as these subsidiary circuits are in-harmonic, or have zero coupling, they will not affect the main oscillation, but small changes in dimensions, or in the method of holding the crystal, or a temperature change, may bring one or other of these circuits into effect.

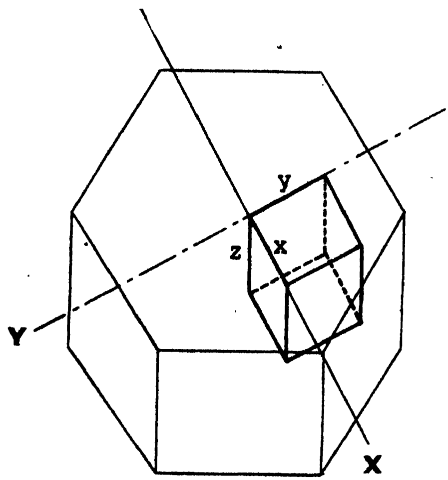
As a plate is ground it is often possible to anticipate the step frequencies, and in the event of a step probably appearing at or near the required frequency, either the crystal must be utilised for a different frequency, or its outside dimensions varied so as to decouple, or change the relative frequencies of edge to thickness. Avoidance to the proximity of stepping points is also desirable on account of the fact that a crystal is usually less active in such a condition.

Crystals of Low Temperature-Coefficient. It is clearly an advantage if a crystal can be cut to have good piezo-electric properties and at the same time to have a low temperature-coefficient. There are many ways in which this object can be achieved, for a limited temperature range.

Since it is found that a Y cut crystal has a positive temperature-coefficient and an X cut crystal a negative

one, crystals having a cylindrical or nearly cubic shape should have a coefficient which approximates to zero. This is found to be so. Thus Marrison produced low-temperature crystals by making them in the form of a thick ring the inner face of which was doubly tapered to a narrower neck at the ring centre.

The low temperature-coefficient is probably due to the general increase of thickness compared with the transverse dimensions rather than by virtue of the ring construction,



z = y
x = 1·19y

FIGURE 192.

although the particular shape enabled Marrison to support the crystal by a line contact giving the least possible mechanical restraint.

T. D. Parkin has produced crystals which are nearly of cubic shape, cut to a dimension which compensates for the difference of temperature-coefficient.

The "Cube" is cut with sides exactly parallel to the X, Y, and Z axes, the length along the X axis being approximately 1·19 times that along the Y, and the depth of the crystal along the Z axis being equal to that along the X axis (see Fig. 192). Not only does such a cut give a very low temperature-

coefficient over a certain temperature range, but the dimensioning is such that the two modes of oscillation coincide and such a crystal is substantially free from stepping troubles.

Since the frequency of a Cubic crystal is given approximately by :

$$f_{\mathrm{kc/s}} = \frac{1910}{y_{mm}}$$

it is seen that the range covered by reasonable size cubes limits their use to the lower end of the high-frequency scale as shown in Table III.

Inclined Angle Cuts. It has already been mentioned that X and Y cut plates suffer from two disadvantages, a bad temperature-coefficient and spurious modes of oscillation. It has been found that plates can be cut at certain angles inclined to one or more of the axes which will be free from one, or other, or both these disadvantages, but it will be best to treat the two cases separately.

The resonant frequency of a crystal has been shown to be due to mechanical stationary waves set up within the quartz and from the formula on page 320 we can write

$$f_{\mathrm{kc/s}} = \frac{1}{2y}\sqrt{\frac{E}{\rho}} \quad . \quad . \quad . \quad (5)$$

where E is the elastic modulus in the plane being considered.

If a quartz plate is changed in temperature, the frequency resulting will be dependent upon how the various terms in (5) vary, and it can be shown that whereas the temperature-coefficient of ρ is independent of orientation of the plate, that of the thickness dimension and of elastic constant vary with plate dimensions and orientation. The chief change brought about by change of orientation is as a result of the alteration in the elastic constant. Since quartz is non-isotropic, the elastic constant varies with direction, and in Fig. 193 is shown a three-dimensional diagram which indicates in polar form the value of E in the various directions, each alternate section considered as a polar curve depicting the change of E along the YZ and XZ planes. Thus the section lying in the plane of the paper is a YZ curve, the edge next adjacent forward is in an XZ plane, that forward again a YZ plane, and the

section seen edgewise is in an XZ plane. From this it is seen that whereas the polar curve is symmetrical about an XZ axis plane (although not circular), it is very asymmetrical about a ZY axis plane, E reaching a value of about 13×10^{11} at 50° on one side, towards any m' face, and a minimum value of 7×10^{11} at 70° on the other. Viewing the crystal from different planes, this asymmetry reverses in direction at each adjacent Y axis.

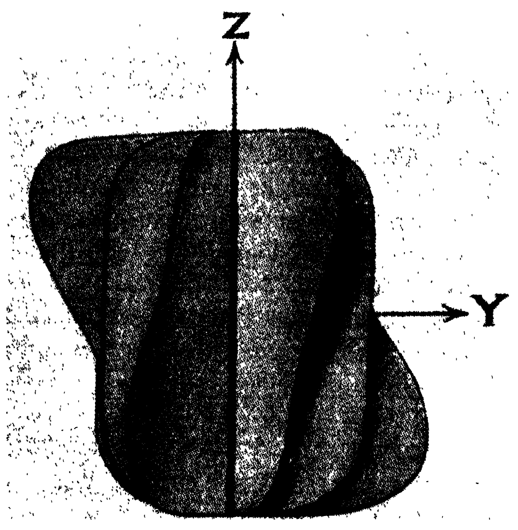


FIGURE 193.

Thus Fig. 188 shows the polar curve in a YZ plane related to a left-hand and right-hand crystal respectively, but facing any positive X axis of the crystal, as seen by the plan and elevation above. It should be clearly realised, however, that since a rotation through each 60° of the crystal before the observer reverses the direction of the sense of the axes, and changes the relative positions of $m\,m'$ faces, it will also reverse the asymmetry of the elastic modulus figure. Thus the clockwise, or anti-clockwise bias of these figures must not be associated with

left or right-handed crystals but only with a change of viewpoint and the relative positions of the *m* and *m′* face.

At and near the *Y* axis plane, that is in the region of greatest asymmetry, the modulus changes rapidly, and it is found that a large variety of cuts inclined at different angles with respect to the *XZ* plane have a substantially zero temperature-coefficient, over a limited temperature range.

Consider a rectangular section of quartz crystal cut from a *Z* cut slat as shown in Fig. 194 (front section removed), and imagine a *Y* cut parallelepiped, whose dimensions are indicated

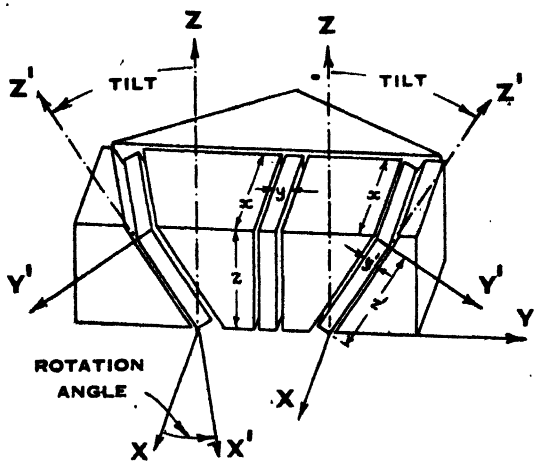


FIGURE 194.

by x, y, z, parallel to the X, Y, and Z axes. As explained previously, such a slice will have an active shear vibration, and a less pronounced edge vibration. If we rotate the cut of such a plate, either way, about its bottom edge, i.e. about the X axis Fig. 194 (right), the more it is rotated towards the horizontal position, the less piezo-electric activity it will have, the relationship of angle of rotation to activity being as shown in Fig. 195: owing to the asymmetric nature of the crystal structure, however, the temperature-coefficient tilt-angle curve for the shear mode assumes a form as shown in Fig. 196, full line. From this we observe that an angular rotation of 35° towards an *m* face or a reversed rotation of 49° towards an *m′* face results in a zero temperature-coefficient being obtained, both types of

vibration being still of the shear type, and both being liable to spurious oscillations, unless precautions are taken to avoid them. The 35° angle cut will have the greater activity, and it is desirable to point out that the sense of angles will change with the hand of the crystal and with the viewing position. By working with an *m* face as guide, however, the correct sense will always be obtained.

Crystals cut in this way, which were first investigated by Bechmann and Koga, are known as *r* cuts by Koga, and *AT* and *BT* cuts by the Bell Telephone Laboratories.

With a tilted plate, since the crystal characteristics are changed, it is usual to denote dimensions x', y', z', and the new axes X', Y', Z', and thus with a plate tilted through 90°

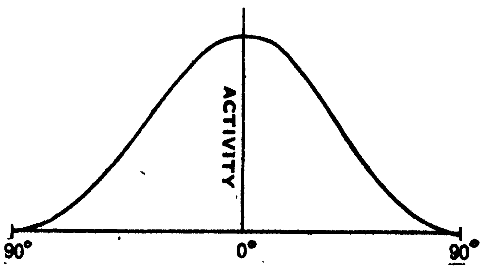


FIGURE 195.

we get the anomalous conditions of a Y' axis coinciding with a Z axis, and a y' side coinciding with a z side, etc.

A simple inclined cut, such as we have described, can clearly be specified by two angles, one giving the inclination from a datum plane, either horizontal or vertical, and the other giving the angle of rotation about the Z axis from a given plane. Because of the asymmetry of crystal properties, it will be necessary for sense to be specified, either in terms of the vertical or horizontal angle.

If, in addition to tilting the plate, we change its horizontal orientation by a bodily rotation about the Z axis (which will not change the angle of tilt), we get what is called by the Radio Corporation of America a *VW* cut, Fig. 194 left, then for any given rotation angle *W* up to some 15° either side of the *XZ* plane it is possible to find a correct tilt angle *V* to give a zero temperature-coefficient. With all such inclined cuts it is

assumed the final plate is cut from the inclined slice with sides parallel and normal to the top and bottom edges of the slice from which it was cut.

If the parallelepiped is cut from an inclined slice with its sides at an angle to the edges of the slice, as shown in Fig. 197 bottom left, such a cut will be a cut "skewed" to all three axes. We could of course imagine such a plate as having been produced by giving to an original Y cut parallelepiped first an inclination about X, then a bodily rotation about Z, and then

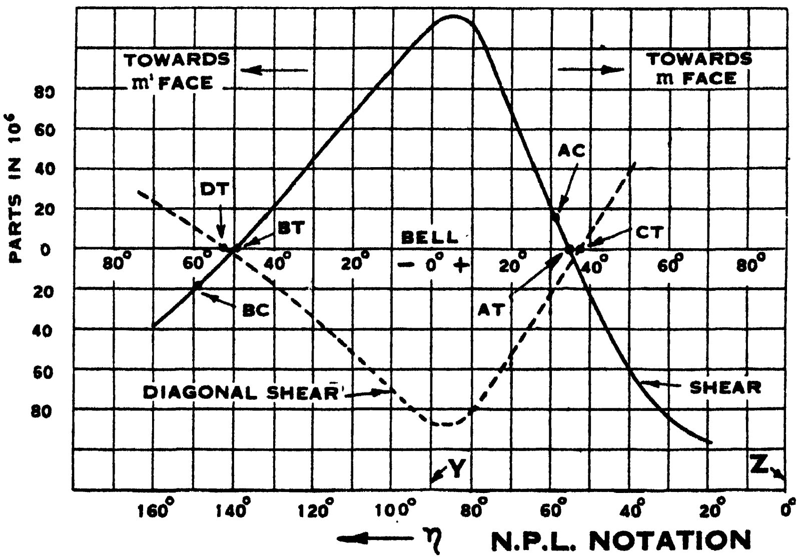


FIGURE 196.

a tilt on to one corner, keeping its major face in the same plane.

The specification of a tilted skewed cut is rather more difficult. Its three axes can be specified in terms of three angles relative to arbitrary planes; or the inclined slice from which the plate is finally cut can be specified in terms of two simple angles as stated, and the inclination defined separately.

N.P.L. Notation. The N.P.L. use as a datum plane for the specification of all cuts, a plane at right angles to the Z axis, called the Z plane. Thus in Fig. 197 x, y, z is a Z cut parallelepiped cut at right angles to the Z axis (shown outside the slab). Any other cut is obtained by first giving to the cutting plane,

now horizontal, a rotation about the X axis of $\eta°$, measured from $+Y$ (that is from the axis emerging from an m face), and then a bodily rotation of $\zeta°$ about the Z axis, the positive angle of rotation being from a positive X to a positive Y axis, the inclined cut being then simply defined as $\eta\ \zeta$.

With such a notation, the rotation angles are without the necessity of sense, positive or negative, because the vertical angle may have any value between 0° and 180°, and thus the asymmetry of the crystal is allowed for by whether the angle is less or greater than 90°. Observe that the vertical angle is

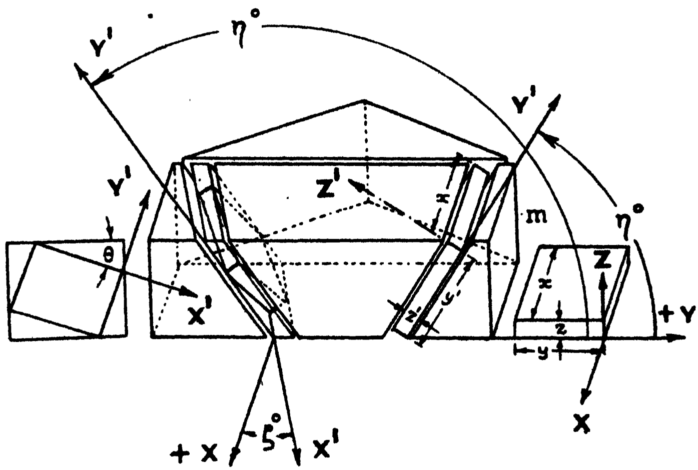


FIGURE 197.

always measured from a positive Y axis, that is from one emerging from an m face. This may be a clockwise or anti-clockwise direction depending not so much upon the hand of the crystal as the viewing position, as previously mentioned. The horizontal angles may have any value up to 60° as rotation though this angle brings the cut back to a similar axis. Thus using the N.P.L. notation, a Y cut crystal becomes 90°, 0°, and an X cut 90°, 30°, the relationship of the notation to this and other cuts being shown in Table 2, and Fig. 196.

In the case of a tilted skew cut, the angle of tilt of the final plate cut from the skewed slice is defined separately, namely as the acute angle (measured counter-clockwise) which one

side of the plate makes with the top edge of the inclined slice, the inclined slice from which the plate is cut having been defined separately as above.

TABLE II.

Alphabetic	N.P.L.		Bell		R.C.A.	
	ANGLE OF					
	I.	*R.*	*I.*	*R.*	*R.*	*I.*
	η	ζ	φ	θ	*W.*	*V.*
					B.	*A.*
Z	0°		90°		90°	0°
X	90°	30°	0°	60°	0°	0°
Y	90°	0°	0°	30°	0°	30°
AT	55°	0°	35°	90°	+35°	30°
BT	138°	0°	48°	30°	48°	30°
CT	39°	0°	39°	90°	−39°	30°
DT	145°	0°	55°	30°	55°	30°
AC	59°	0°	31°	90°	−31°	30°
BC	149°	0°	50°	30°	59°	30°
GT	51°	0°	45°	—	—	—

NOTE.—In the above table *I* shows the angle of inclination and *R* the rotation angle.

American Notation. In America two systems of notation have arisen, one originated by the Radio Corporation of America and the other by the Bell Telephone Laboratories. Both differ from the N.P.L. by adopting the *Z* axis as datum for the angle of inclination, and not the *Z* plane, and they differ from each other and from the N.P.L. by a different convention as regards the horizontal angle of rotation.

The Bell Telephone Laboratories identify their main types of crystal plates by letter notation and an inclination angle positive or negative, depending upon whether the new Z' axis is towards or away from an *m* face as indicated in Fig. 196. Whereas the inclination angles for the various alphabetic cuts are shown either side of the *Z* axis, their specification is, however, determined by an angular deviation from the three axis planes. Thus in Fig. 198, which shows a rectangular plate inclined to all three axes, its orientation is defined by the Bell Laboratories in terms of three angles ϕ, θ, . Thus although in their alphabetic designation of *AT*, *BT*, etc., plates the new axis Z' is defined towards or away from an *m* face, in their

specification the inclination angle is defined by $\phi°$ without sense. The asymmetry of the crystal is therefore allowed for by giving sense to the rotation angle θ measured from an edge not a face, a positive sense being anti-clockwise, and the negative sense clockwise, for a right-handed crystal. As mentioned previously the Bell Telephone Laboratories definition of right-handed and left-handed quartz is opposite to that of other workers in the piezo-electric field.

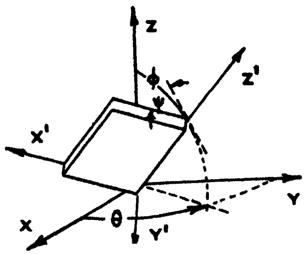


FIGURE 198.

The definition of the final tilt is obtained by the angle $\psi°$ which is defined as the angle by which a cut inclined $\phi°$ and rotated $\theta°$ about a horizontal axis, is then rotated about the new Z' axis. This second tilt will clearly change the direction of the X' and Y' axes from the direction before tilting, and it would appear therefore that this type of specification is somewhat difficult of interpretation.

The R.C.A. have a different system of notation. For the angle of inclination, they adopt the Z axis as datum and define the angle of inclination as $B°$, positive *away* from the m face, and negative towards it. For the rotation angle about Z, the specification is somewhat involved, but virtually it may be considered as specifying the angle (measured clockwise or anti-clockwise) from an equivalent Y axis. Thus a Y cut plate will be defined as A 30° B 0°.

Of these notations, correlated in Table 2, the N.P.L. system is adequate for all types of cut, it is quite without ambiguity, and is the only system which defines a plate inclined to all three axes in a simple manner. Where no ambiguity exists, the designation of plates by letter appears highly desirable, such for instance as "X" plate, "Y" plate, etc.

The η 55° ζ 0° (AT) and the η 138° ζ 0° (BT) zero temperature cuts previously mentioned are thickness shear modes, suitable for high frequency work as indicated in Table III. Since, however, a crystal has a twofold dimension at right angles, it would be expected that two cuts complementary to 55° and 138° would also give a zero temperature-coefficient, but having a different mode of oscillation—namely, across

the face of the plate. This is found to be so, and low-frequency, diagonal-shear-mode plates at values near η 145° (*DT*) and ζ 51° (*CT*) inclination are found to have a zero temperature-coefficient, the first being complementary to the *AT* plate and the second to the *BT*. The relationship of the *AT* and *DT* plates is shown in Fig. 199, which also indicates the modes of vibration, and the dotted curve in Fig. 196 shows the relationship of temperature - coefficient and angles of inclination of diagonal-shear plates.

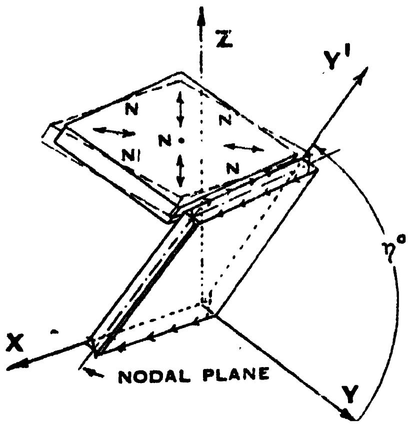


FIGURE 199.

All the plates we have mentioned suffer to some extent from coupling effects to other modes and as mentioned previously the "temperature-frequency change" relationship does not hold over an infinite temperature range, but is limited. In most cases the shape

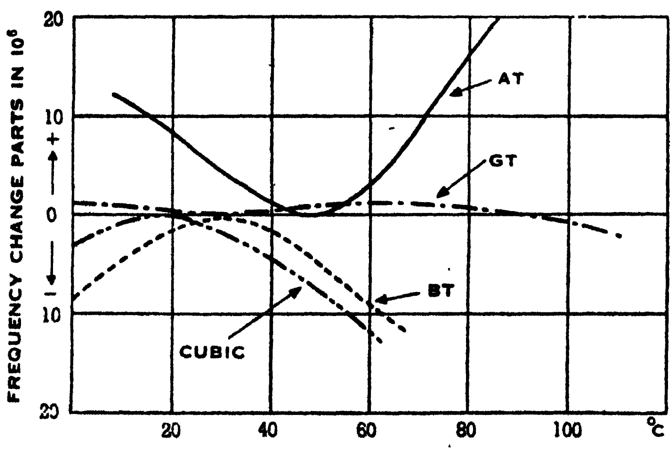


FIGURE 200.

of the "temperature-frequency change" curve is parabolic as shown in Fig. 200, where curves for selected crystal plates are given.

The Radio Corporation of America have found that with a change of rotation angle away from the X axis, and a corresponding alteration of the inclination angle, a whole series of cuts, with appropriate angular notations (designated VW), can be found with an extended range of temperatures over which the coefficient is zero.

Zero-Coupling Crystal Cuts. It is found that a simple inclined cut, oscillating in h.f. shear mode, can be found where the coupling between edge and thickness modes is zero. Two such cuts are η 59°, ζ 0°, and η 149° ζ 0°, so called AC and BC by the Bell Telephone Laboratories, but by a change of rotation angle ζ, a series of cuts having similar characteristics will be found.

The Bell Telephone Laboratories have also devised a cut having a limited frequency range, which is most interesting as it not only has zero coupling, but a zero temperature-coefficient over a very wide temperature range. This cut is called by them GT, and in N.P.L. notation is a rotated inclined cut of η 51°, ζ 0°, θ 45°, and is obtained by first cutting a CT η 51°, ζ 0°, and then cutting from this a crystal plate with sides at 45° to the edges of the block. It has relative edge dimensions of width to length of ·86 to 1, and the width determines the frequency obtained.

According to the Bell Laboratories such a crystal oscillates in compression and extension about a nodal line across the centre of the plate. This can be seen from Fig. 199, where the top plate shows the diagonal shear vibration of a CT type plate, and it will be clear that if a plate was cut from this with sides at 45°, there will be extension and compression about a diagonal of the CT cut plate. The extended temperature range over which its frequency remains constant can be seen from Fig. 200.

The overall dimensions of the various plates we have mentioned are governed partly by power considerations and the use to which the crystal is to be put, partly by holder design and to avoid spurious oscillation modes, but for thin plates a face dimension of about 25 mm. square appears to be a general dimension, the actual thickness being usually determined by the frequency required. The table following gives relevant information regarding the various types of cut we have enumerated.

Table III.

Cut.	Type of Oscillation.	Frequency kc/s given by :— (x, y, in mm.)	Freq. Range kc/s.	Frequency Change. Parts in 10^6 per 1° c.	Remarks.
X	Compression and Extension	$\frac{2750}{y \text{ and/or } x}$	40— 1000	—20 to —50	Active. Nearly obsolete for transmitters. Steps.
Y	Shear	$\frac{2070}{y}$	500— 3000	+66	Very active. Nearly obsolete for transmitters. Steps.
AT	Shear	$\frac{1630}{y'}$	500— 2000	Zero at correct temperature	Very active. In great use. Steps. Can be clamped.
BT	Shear	$\frac{2500}{y'}$	2000—15000	,,	Active. In great use. Steps. Can be clamped.
CT	Diagonal Shear	$\frac{3100}{x}$	100— 200	,,	Very active. Can be only clamped at centre.
DT	Diagonal Shear	$\frac{2100}{x}$	70— 150	,,	Very active. Can be clamped at centre.
GT	Compression and Extension	$\frac{3292}{x'}$	60— 1000	Zero over wide range.	Active. Clamp on centre.
AC	Shear	$\frac{1620}{y'}$	500— 2000	+20	Active. No coupling.
BC	Shear	$\frac{2560}{y'}$	2000—15000	—20	Active. No coupling.
Cube	Shear	$\frac{1925}{y}$	75— 750	Zero at correct temperature.	Active. Free of steps. Clamp at critical points.
VW	Shear	$\frac{1600}{y'}$ to	500— 2000	,,	Very active. Modification of AT and BT.
VW	Shear	$\frac{2500}{y'}$	2000—15000	,,	

Resume of the Various Cuts. From a study of the above table it is clear that by the selection of the proper cut, a great range of fundamental frequencies can be covered.

Not all these crystals are used, however, for oscillator work, as they are not active enough and would therefore be difficult to start into oscillation and probably give only very small output.

The most active plate is the *Y* cut, but its bad temperature-coefficient precludes its use in most cases, and if we exclude the *X* cut crystal for the same reason we can assume that of the others, the *AT*, *BT*, *CT*, *VW*, *DT*, *GT*, and Cubic are useful for oscillators. The *X*, *AC*, *BC*, and *GT* are used as resonators, and as impedances in filter networks.

In its range the Cubic crystal is the most economical as regards quartz and manufacturing costs. Not only is it the smallest in overall bulk, but because the faces are parallel to the principal axes of the crystal, it is easy to set up on the jig and cut almost straight away to finished dimensions. Further because of its shape a much greater percentage of the good quartz may be utilised. This type of crystal is free from spurious oscillations, but its temperature-frequency characteristic is parabolic. It is not critical as regards holder design and its small bulk makes for a compact unit.

Of the inclined cuts, the *AT* cut is most active, it can be made for a large frequency range, and it gives a good output and can be clamped in its holder. It suffers, however, from stepping troubles and like all angle cuts it is wasteful of material in manufacture. The *BT* cut can be made to higher frequencies than any other type of crystal, it can be clamped, but it is less active than the *AT* cut and gives smaller output. The *CT* and *DT* cuts cover a lower frequency range and are alternative to the Cubic type of crystal. They require careful clamping at the centre thus requiring a special type of holder, and they cannot be overloaded. One of the great advantages is, however, that they can be ground to a definite frequency at a definite temperature without great difficulty. This is possible because the frequency of the plate can not only be decreased by grinding, but also increased as well. Increase of frequency is obtained by grinding the corners away, and decrease of frequency is possible by grinding away the centre, and thus a correction for over-

grinding can be made without difficulty, although it will be realised that the temperature-frequency curve is parabolic.

The *GT* plate appears to be suitable for all classes of work and its mode of oscillation lends itself to rigid clamping by a holder of simple design.

Crystal Holders. It will be obvious that the crystal holder will play some part in determining the final frequency of the crystal unit as a whole. Although the crystal itself is the main factor for determining the frequency of oscillation, the method of mounting must influence the frequency, even though this will be a second order effect, as the holder is bound to add capacitance and damping. Thus, although a simple and inexpensive type of holder can be devised for crystals having a wide tolerance, such for instance, as used in ship stations, portable sets, and rough-check wavemeters, where a precision of not more than 1 part in 20,000 may be required ; the design of a holder for crystals used for high precision work (such, for instance, as in a frequency-checking station) will need the most careful consideration.

Before modern methods of grinding and surfacing to a precise value were developed, it was common practice to carry the crystal in a holder containing an air gap, which could be adjusted to bring the crystal to the required frequency, a variation of 1 part in 2,000 being easily possible by gap variation with the ordinary *Y* cut crystal, a typical variable-gap holder being as shown in Fig. 201. This is not good practice where only a small tolerance is possible, and present-day holders are usually designed either without gaps, or when a gap is used, it will be fixed, or possibly given the very smallest amount of adjustment. For the highest precision, the holder will often be evacuated, or partially evacuated, as this protects the crystal from moisture, barometric changes have no influence, and the vibration of the crystal cannot set up air resonance changes, but generally speaking such crystals will not be used for oscillator work.

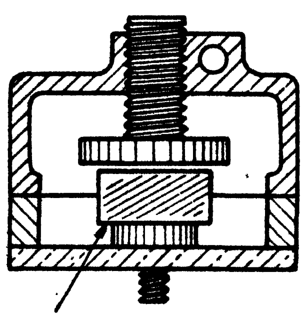


FIGURE 201.

Holders generally are of two main types. One which clamps the crystal either on the faces of the crystal, where this is possible, or nodally, on point, or line contacts depending upon

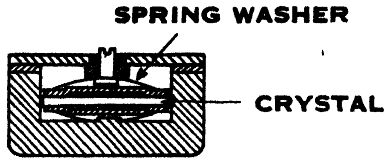


FIGURE 202.

the type of crystal being held. Secondly a type of holder having a fixed gap (or possibly having a very small variation) and designed to carry the crystal resting on one face and lightly constrained laterally. With the Cubic crystal another type of holder is sometimes used in which the crystal is suspended in air between the electrode faces.

A simple type of holder, in which the crystal is clamped lightly between two electrodes by means of contact springs is

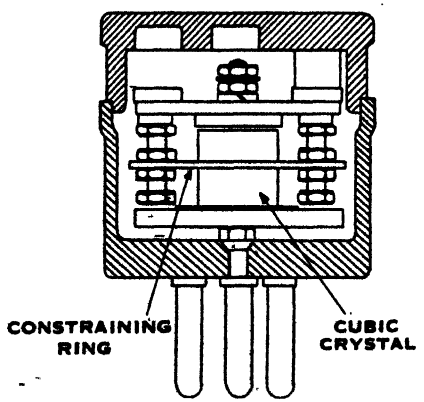


FIGURE 203.

shown in Fig. 202, such a holder being useful for oscillator crystals, using an *AT* or *Y* cut.

Figures 203 and 204 show two types of holders designed for Cubic cut crystals. In the first a small air gap is allowed between the top of the crystal and the electrode, and the crystal is held laterally inside a thin ring, notched to take the corners

of the crystal. Such a crystal unit would be suitable for tolerances up to 2 parts in 10^6, and would be used for oscillators. The holder shown in Fig. 204 is novel as the crystal is suspended in a cradle of silk threads, with threads stretched across the faces of the crystal to prevent sideways movement, the crystal and cradle assembly being mounted centrally between the electrode faces so as to leave a small air gap each side. Such a crystal assembly has a precision within 3 parts in 10^7 and would be of use for frequency-checking equipment.

For precise work with certain of the inclined cuts, clamping can be resorted to, but only at selected nodal points. It is clearly a great advantage if the design can be such that firm

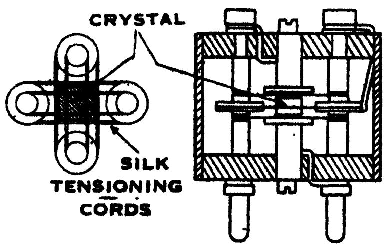


FIGURE 204.

clamping is possible, as it solves the transport problem very considerably, for with unclamped crystals any mechanical disturbance tends to shift the frequency. For instance, in the *GT* cut, where the crystal vibrates about a nodal line across the centre, a holder is used which grips the crystal between two wedge-shaped jaws along the nodal line. This is clearly a very simple type of holder to design. Another method of clamping is between hardened pointed steel pins, nodal points being chosen, which may be on the face as in the case of the *CT* and *DT* cuts, or along the edges as in the case of *BT* cuts.

The great variety of cuts which appear to be possible having either a zero temperature-coefficient, or small coupling between modes, or both, is so large, that the technique of crystal production and holder design is clearly only now in an early stage of development, and we should see considerable advances in the next few years.

Circuit of the Valve Maintained Quartz Crystal. A very good type of circuit is that due to Pierce and Miller and shown in Fig. 205, a modification of which consists of placing the crystal between grid and filament. The method of maintaining oscillations may be explained simply as follows :

The crystal when vibrating is very similar in its effects to a parallel-resonant circuit having one end connected to the anode and the other to the grid. The oscillatory voltages applied to anode and grid are, therefore, opposite in phase and conditions are correct for the maintenance of oscillations. Switching on the circuit will be sufficient to shock the crystal into oscillation, and this will then be maintained with energy from the anode battery supplied in the correct phase through the valve.

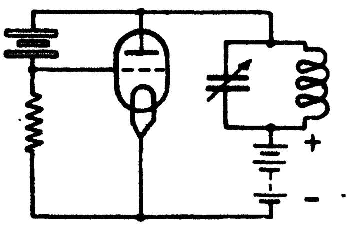


FIGURE 205.

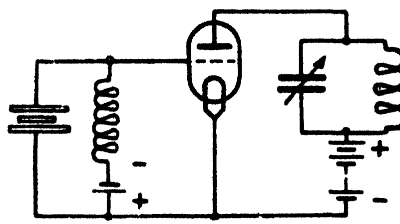


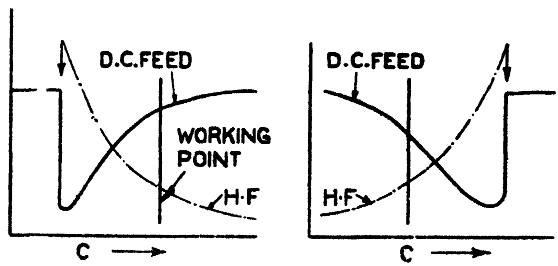
FIGURE 206.

It is evidently essential that the anode circuit should have a high impedance in order that the anode may not be shorted to the filament as far as high frequency currents are concerned. This high impedance may be provided by an ordinary resistance, but a more efficient circuit is usually produced if a parallel circuit nearly in tune with the crystal is used. As the circuit condenser is reduced from a very large value, oscillations build up until a maximum is reached. Further reduction of condenser value produces a sudden fall in output until oscillations cease just before the circuit comes into tune with the crystal frequency. It is necessary, therefore, that the resonant frequency of the anode circuit should be below that of the crystal, in the case shown in Fig. 205.

The alternative circuit of Fig. 206, suggested by Miller, is also much used. In this case the crystal provides the equivalent of a tuned-grid circuit and the necessary " back

coupling" is obtained through the valve capacity. In order to have the relative phases correct for the maintenance of oscillations it is necessary that the anode circuit be detuned above the crystal frequency.

With both types of circuits care must be taken that the natural frequency of the resonant circuit is not too near that of the crystal, in which case electrical self-oscillation may ensue. Although the effect of this is to increase the high frequency output, the resonant circuit tends to "pull" the crystal, and in such a case the crystal is not performing its proper functions of driving. Fig. 207 showing H.F. output with tuning, curve "a" being for the grid-anode circuit, curve "b" for the circuit of Fig. 206 and for the grid-



a FIGURE 207. *b*

filament circuit. As a general rule the grid-anode circuit will be used when great precision is required, and the second circuit when high output is necessary and precision not so important.

Bridge-Stabilised Oscillator.[12] A crystal circuit for high-precision work, suggested by Meacham, is the bridge-stabilised oscillator, in which the amplitude of oscillation, as well as the frequency, is automatically kept constant by a special bridge.

The circuit is shown in Fig. 208, where R_1 is a resistance the value of which depends upon its temperature (usually a tungsten-filament lamp of low wattage) R_2 and R_3 are normal resistance elements, and C the crystal arm. One diagonal of the bridge is connected across the input of an amplifier, and the output of the amplifier connected back across the other diagonal of the bridge.

If the bridge was in exact balance, no oscillations could occur, as there would then be no transfer of energy between the opposite diagonals of the bridge, but the introduction of a resistance which varies in value as the current passing changes, allows the circuit to oscillate between fine limits of amplitude.

The action is as follows: Before switching on the amplifier, the resistance of the cold tungsten filament is low, the bridge is right out of balance, and oscillations therefore build up; but the rising oscillator current through the bridge arm heats up the filament and its resistance value approaches the balance conditions until the attenuation thereby introduced equals the gain, when an equilibrium condition is arrived at. The fine control exerted by the crystal is dependent upon the fact that

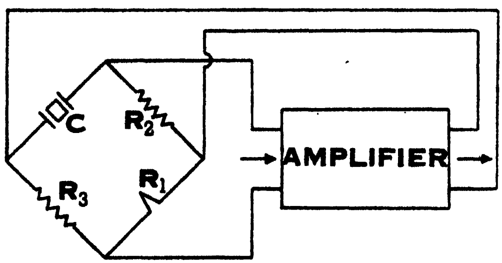


FIGURE 208

the phase shift through the bridge must be equal and opposite to that through the amplifier and the crystal thus impresses its frequency on the combined unit.

The circuit is capable of providing an output constant in amplitude and frequency and of sinusoidal waveform, in spite of power-supply fluctuations and changes in circuit elements, and a frequency-constancy equal to, or better than, 2 parts in 10^8 is possible.

Oscillators with Electrical Resonant Circuits. The reader is assumed to be familiar with the principle of the valve *LC*-circuit self-oscillator and we shall confine ourselves here to modifications necessary at the higher radio frequencies and to a brief study of the factors upon which stability of frequency depend.

The separate reaction-coil circuit frequently employed at low radio frequencies is not suitable at higher frequencies

because the voltage applied to the grid is not in exact phase opposition to the anode voltage, as it should be.

The Hartley or Colpitts circuits, Fig. 209 and 210, may be employed, but it is important to note that the valve capacitances

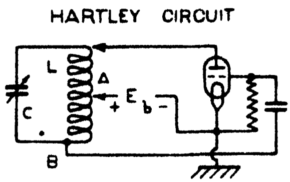


FIGURE 209.

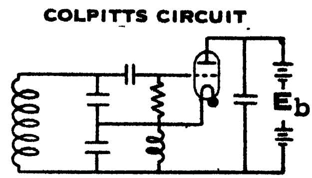


FIGURE 210.

now have an important effect upon the reaction conditions. The grid anode valve capacity may have a sufficiently low reactance to provide all the reaction necessary and hence the circuit shown in Fig. 211 (due to Franklin) is frequently employed for short waves. The condenser C_b is made sufficiently large to constitute a short circuit as far as the oscillation frequency is concerned. The circuit diagram therefore appears to be very similar to that of the Hartley circuit, but in practice the coils are usually completely decoupled and reaction is chiefly due to valve capacitance. The chokes isolate the oscillatory circuit from other circuits, filament chokes being only necessary

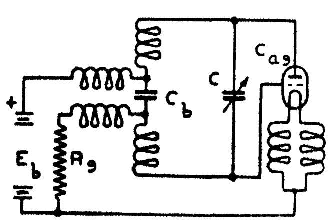


FIGURE 211.

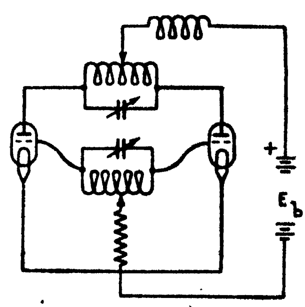


FIGURE 212.

when frequencies above about 30 Mc/s are to be produced. Where filament chokes are used they need careful design so as to build up the correct voltage conditions at the filament.

Push-Pull arrangements are very suitable and Fig. 212 shows a circuit similar to that of an amplifier but without the

neutralising condensers. A very important difference between this circuit and that of the amplifier is that in the oscillator the grid circuit should be detuned so that its own resonant frequency is about 1·25 times that of the anode circuit, which should control the frequency produced. Very inefficient conditions occur if the grid circuit is also tuned to the oscillation frequency.

Present-day short wave commercial transmitters do not employ a self-oscillator coupled directly to the aerial because the frequency stability is too poor. The production of appreciable power by a self-oscillator is, therefore, seldom attempted and questions of power efficiency are therefore quite subsidiary to the main problem of providing a correct and unvarying frequency which can be amplified as necessary.

The production of a constant frequency is also important in reception because most receivers will contain an oscillator which should be as stable as possible but which will usually have to be simple and capable of tuning quickly over a range of frequencies.

Requirements for Constant Frequency. As previously mentioned the great success of the piezo-electric oscillator is due in the main to two factors: zero temperature-coefficient of the oscillating element, and its inherently large Q value. If we are to use an LC circuit as the controlling element it should, therefore, have the same qualities.

This means, in the case of a tuned circuit, that the product of inductance and capacity should remain constant under all circumstances. This can either be achieved by making each unit have a zero temperature-coefficient, or compensating for any changes which do take place. Variations of inductance and capacity will occur if coils and condensers change in size with temperature, or if their parts move due to poor mechanical design or excessive vibration.

The resonant frequency of an LC circuit, in which no precautions have been taken, may vary by as much as 1 part in 1,000 for quite ordinary temperature changes, and where vibration is considerable, as on ships and aircraft, carefully designed suspensions may become necessary for units containing oscillators.

Coils which show a small variation of inductance with

temperature may be made by depositing the conductor on to a former of suitable material. In this way the expansion of the former becomes the controlling factor and some suitable ceramic materials have much smaller coefficient of expansion than metals. Coils have also been designed in which the effect of the axial expansion (which reduces the inductance because it reduces the mutual inductance between adjacent turns) balances the effect of the radial expansion (which increases the inductance). The variation of self-capacity also produced by a change of temperature complicates the problem.

In the design of variable air-condensers to have small temperature-coefficients, the most important thing is to prevent relative axial movement between the fixed and moving vanes since a small movement will make a relatively large change of capacity. Some of the condensers employing a ceramic dielectric (for example, "Calit") have a small temperature-coefficient.

Apart from making the coil and condenser as little dependent upon temperature change as possible, we can also arrange that an increase of temperature increases the inductance but reduces the capacity, thus keeping their product constant.

Whilst the values of coil and condensers are the main control of the frequency yet other factors exert some influence, and in designing a constant-frequency oscillator these "second order" effects will be of great importance.

It is first necessary to realise that the frequency generated will not usually be exactly that of the resonant frequency of the circuit. This may be simply explained by reference to the circuit of Fig. 213, and the corresponding vector diagram, but the same conditions may occur with any of the circuits usually employed for oscillators. The condenser is considered to be free from loss and all the circuit resistance located in the coil. The voltage applied to the grid will be 90° out of phase with the current in L, if alternating grid-current is neglected. But I_a is not in phase with E_{LC} and hence the oscillating frequency cannot be the resonant frequency, but is actually somewhat higher. It needs to be emphasised that a valve self-oscillator tends to adjust itself to working conditions similar to those of a Class C amplifier, running into grid current and being limited thereby. It is the curvature of the valve characteristic which

puts a limit to the growth of the oscillation and therefore determines its amplitude. Explanations of valve oscillator operation, therefore, which depend upon vector diagrams or vector algebra cannot give quantative results and are only very rough approximations.

The valve capacitance also forms part of the oscillatory circuit and thus changes of valve capacitance will affect the resultant frequency unless precautions are taken, either to compensate for them, or reduce their effect.

We have now to consider the effect which changes in the maintaining system will have upon the frequency. The effective inter-electrode capacities of the valve depend to a

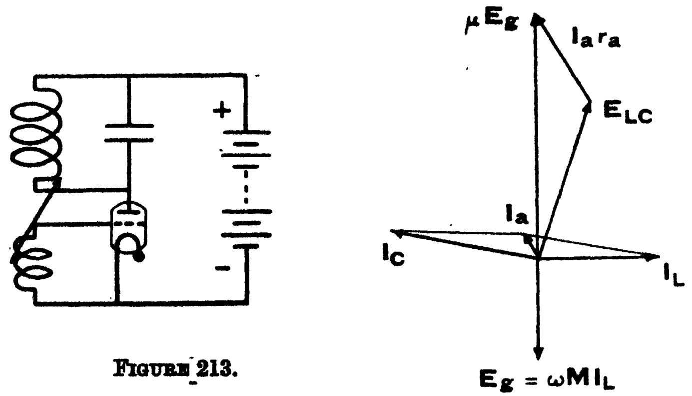


FIGURE 213.

small extent upon the voltages on the electrodes, and since these capacities are across the resonant circuit, this means that the frequency generated will depend upon the supply voltages. For this reason the desired high Q value should not be obtained by using a high L/C ratio because this enhances the effect of the shunt valve-capacities and of any change in load reactance. The resistance of the valve also depends upon the supply voltages and this resistance affects the phase angle between E_{LC} and I_a, so that this is another link between frequency generated and supply voltage. The use of a valve having a high value of r_a reduces variations due to this cause.

The frequency stability is also improved if oscillations are limited in amplitude and this may be done by inserting a grid resistance in such a way that the grid only becomes the smallest amount positive.

Evidently, if a circuit has a high Q value, only a small supply of energy is necessary to sustain an oscillation and hence the coupling between maintaining circuit and resonant circuit can be loose, thus reducing the effect of changes in the maintaining system. It is possible to insert reactances in the anode and grid leads of the maintaining valve which have the effect of bringing the oscillating frequency nearer to the resonant frequency of the circuit than it would otherwise be, and hence reducing the effect of changes in the maintaining system.

Franklin Master-Oscillator. An interesting type of precision oscillator is that developed by Franklin and Witt, the circuit diagram of which is shown in Fig. 214, and the mechanical arrangement in Fig. 216.

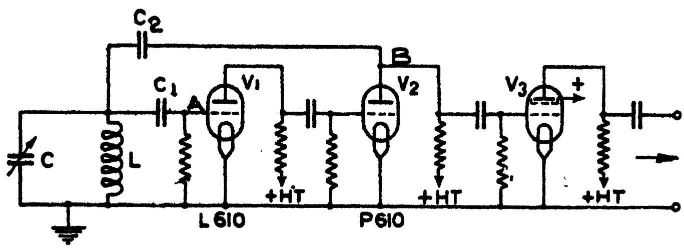


FIGURE 214.

Associated with the resonant circuit are two valves V_1 and V_2. Two valves are employed partly because the higher gain thereby obtained enables the maintaining circuit to be coupled very loosely to the resonant circuit and partly because one end of the resonant circuit may then be earthed, which makes the mechanical design easier. Since the voltage applied to the grid of V_2 is opposite in phase to that applied to V_1, it is evident that the anode of V_2 should be coupled back to the same end of LC as the grid of valve V_1. The coupling capacities C_1 and C_2 are only about $1\,\mu\mu F$ and therefore smaller than the valve capacities, and hence the effect is that of a circuit having low grid and anode taps, as Fig. 215 makes clear.

The oscillator is arranged to work at a lower frequency than that finally required, a usual and desirable feature of most constant-frequency drives, and the valve V_3 is followed by a normal frequency-multiplying amplifier.

The method of self-compensation for temperature of the resonant circuit is of considerable interest, and reference should be made to Fig. 216.

A cylindrical brass case *A*, which is earthed, forms one plate of the main tuning condenser, and acts as a support for the unit as a whole. Concentric with the cylinder is the insulating former *F*, carrying the inductance winding *L*, one end of which is connected to the end of the case, and the other to the insulated condenser plate *B* ; this is of half-tubular form, and is built on the free end of the insulated tube, its outer end being mounted on the insulating diaphragm *C*, which rides on the brass rod *D*.

This rod forms the axis of the cylinder, and is screwed through the cylinder end with a fine thread, the head of the rod

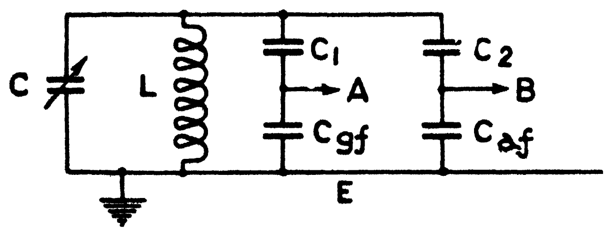


FIGURE 215.

being slotted to permit adjustment of the endwise position of the rod. Part way along the rod is a circular plate *G*, which forms a capacity with the end of the condenser plate *B*, and the relative position of these determines the amount of compensation.

Between the condenser plate *B* and the case is a second pair of circular half-plates *EE* insulated from both, and carried on an insulating end cheek *J*. The position of these plates determines the main circuit capacity, and sufficient variation is provided to give a 10% wavelength change. This adjustment is made by rotation of the plates, operated through a worm wheel.

The operation is as follows : By proper selection of materials the chief endwise expansion is designed to take place in the insulating former, and thus as it warms up, the plate *B* moves away from the compensating plate *G*. The expansion of

the winding causes its inductance to be increased, but this increase is offset by the reduction of capacity between the plate *G* and *B*.

To make the drive immune from external capacity changes and vibration, and to some extent to prevent too sudden a temperature change, the whole unit is slipped inside a second brass cylinder (not shown) between the two being a layer of thick felt. This second cylinder forms a support for the two oscillator valves and the coupling valve, and thus a self-contained compact unit results. The coupling condensers, which

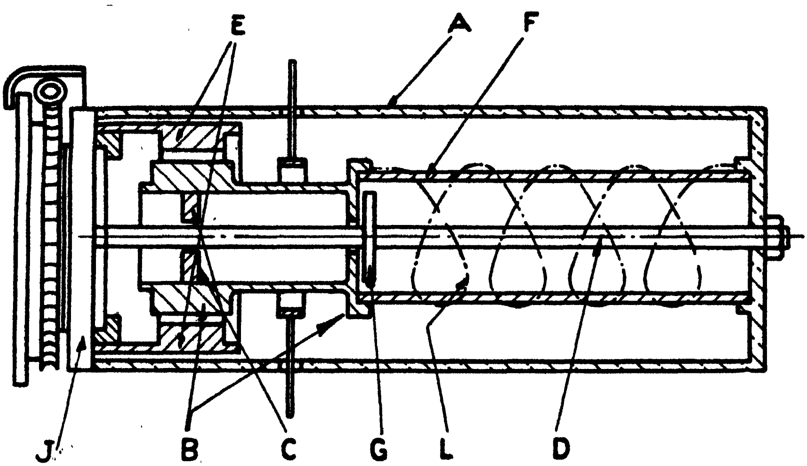


FIGURE 216.

of course must be connected to the free end of the coil, are simply two very minute brass strips held in a mica ring, connections being made by two pins projecting through both cases.

Oscillators using Resonant Lines.[6, 7] The ordinary coil and condenser resonant circuit becomes very small in size and difficult to design when it is to tune to frequencies above 100 Mc/s. Also, self-capacity in the coil and inductance of the condenser leads and plates become so important that it is not even approximately true that the current is the same in all parts of the circuit. At these frequencies it becomes practicable to use arrangements in which *L* and *C* are distributed over the complete circuit, the size then being more convenient and the *Q* value may be much higher. From what we

have discussed in Chapter VI it was seen that a short-circuited line, one-quarter wavelength long, provides the equivalent of a parallel-resonant circuit, since its input impedance is a high resistance. By a suitable design (placing supporting insulators where the voltage is low, etc.) values as high as 10,000 may be obtained for Q at high radio frequencies. If Q is calculated from conductor loss considerations alone, it actually increases with frequency, being proportional to $\sqrt{f}$. At 300 Mc/s such a circuit will be less than 25 cm. long so that it can be

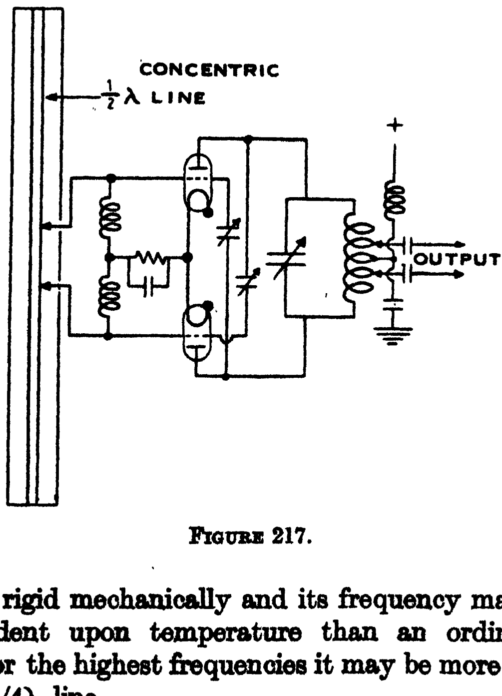


FIGURE 217.

made very rigid mechanically and its frequency may be much less dependent upon temperature than an ordinary tuned circuit. For the highest frequencies it may be more convenient to use a $3/4\lambda$ line.

Evidently a half-wave line, open-circuited at both ends, may also be employed and the frequency varied either by changing the line length, or by placing a variable condenser across the far end. The half-wave line is particularly suited to "push-pull" circuits, an example being shown in Fig. 217, the nearer the tapping points are made to the centre of the line, the nodal point, the less influence the maintaining circuit will have, but the less the output.

Both parallel-wire and concentric-tube lines have been employed, but the latter are now established as the more useful for the purpose, on account of their greater rigidity and the fact that a screened arrangement is produced.

Several methods of compensating for temperature changes have been developed. In one circuit developed by the R.C.A. a pipe containing oil is run inside the outer conductor of the concentric line and therefore takes up the temperature of the line. The expansion of oil resulting from a rise in temperature causes the movement of a condenser plate and thereby compensates for the change in length of the line.

Powerful short-wave transmitters have been constructed in which the oscillator (controlled by a line of the type just discussed) was followed by only one stage of amplification and the frequency stability was found to compare favourably with a crystal-driven transmitter.

Special Types of Resonant Circuit. At the higher frequencies, other special forms of circuit beside the resonant line become mechanically possible, which have the merit of mechanical rigidity and high Q value and an equivalent low L/C ratio. Like the resonant line the inductance and capacity are to a large extent distributed.

The design of such resonators has been developed by Kolster,[8] Hollman and others. A typical design is shown in Fig. 218. From this it is evident that the capacity is mainly distributed between the edges of the two copper shells, whilst the inductance is mainly associated with the central rod. A range of frequencies can conveniently be obtained by sliding one shell along the tube, but a difficulty with these resonators is that it is not possible to couple in the maintaining valve symmetrically and in circuits of large dimensions the resulting current distribution is irregular. The approximate dimensions of a resonator for 60 Mc/s are shown in Fig. 218, the Q value being 1500 and the L/C ratio 300. Resonators which are

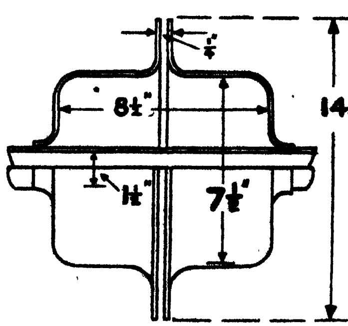
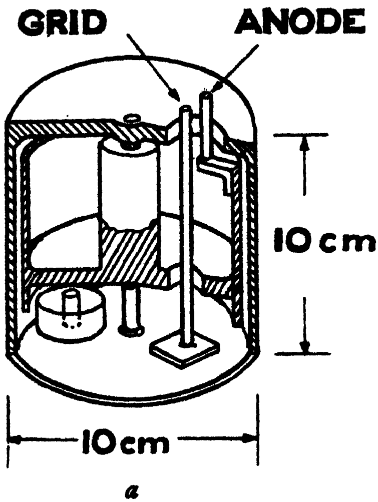


FIGURE 218.

geometrically similar will have frequencies inversely proportional to their linear dimensions.

A somewhat different arrangement, particularly suitable for a small oscillator for measurement purposes, has been developed by the General Radio Company and the circuit is shown in Figs. 219a and b, the first showing the resonant circuit and the second the diagram of connections. The capacity of the circuit is mainly between the inner and outer cylinders and the inductance due to the inner rod, and a circuit having dimensions given in the figure will resonate to about 100 Mc/s,



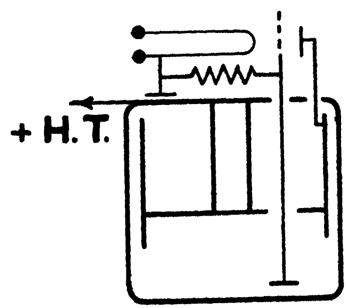


FIGURE 219.

and has a Q value of approximately 2500. Both this and the previous arrangement are self screened. In the latter, the outer cylinder is of brass, and the inner of copper, and the different temperature-coefficients of expansion of these two metals results in the capacitance having a small negative frequency temperature-coefficient, which balances the positive temperature-coefficient of the inductance.

Oscillators for Ultra-high Frequencies. If we raise the oscillation frequency of an ordinary valve in a conventional circuit (including circuits just discussed as conventional) we find the output falling off until eventually oscillations cannot be produced. Using a triode with ordinary 4-pin base, this

limiting frequency might be about 100 Mc/s. The reasons for this are largely the same as were discussed in connection with amplifiers (see page 303). Lead capacities and inductances make it impossible to reduce the effective LC of the resonant circuit beyond a certain minimum and these reactances, together with the effects of the transit-time, make it difficult to obtain proper reaction conditions. It has been shown by approximate theory that if the transit time is 1/12th of the oscillation period, then the efficiency is reduced by at least 10%.

Because of the low efficiencies and other difficulties associated with electron oscillators (see Chapter XII), however, considerable work has recently been directed towards making the conventional reaction-oscillator function at higher frequencies.

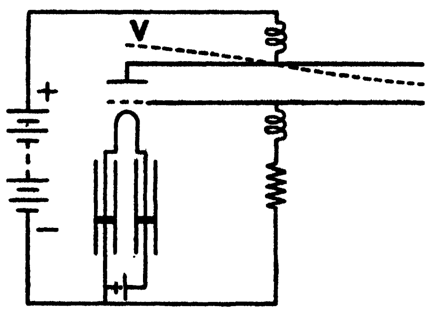


FIGURE 220.

It has been found that by modifying the design of circuit considerably to take account of the limiting factors previously mentioned, and taking particular precautions over the actual valve design, much higher frequencies may now be obtained efficiently.[9]

Appendix V gives particulars of recent valves specially designed to operate on very short wavelengths.

A circuit suitable for an oscillator at such frequencies is shown in Fig. 220. It will be seen that a half-wave line is employed and also the filament is fed through quarter wave lines which present a high impedance at the filament ends and confine the oscillations to the proper circuit. The oscillation frequency is governed by the half-wave line between grid and anode but the filament lines need to be tuned to give appreciable output. The line is not actually $\lambda/2$ long, of course,

because of the electrode capacitance. In fact at the highest frequencies the line may only be of the order of $\lambda/4$ long.

Valves have also been constructed by the G.E.C. laboratories, in which the anode and grid are extended through the glass envelope to form a concentric resonant-line. The low characteristic impedance of the concentric line is an advantage, because a given electrode capacity shunted across the line has a smaller effect in reducing the oscillation frequency if the characteristic impedance is low.

Wavemeters. If interference between stations is to be prevented and the available frequencies used to the maximum extent, transmitters providing frequencies which are both constant and correct are essential.

Since the earliest days of wireless, frequency has been measured by setting up as permanent a resonant circuit as possible, and calibrating it. The simplest method of calibration is to measure the inductance and capacity by low frequency bridge methods and calculate the resonance frequency, but this is quite inadequate for present-day requirements, and calibration must be carried out by reference to a source of accurately known frequency.

The accuracy of a wavemeter depends principally on the mechanical permanence of the coil and variable condenser, and upon the effect of the resonance indicator upon the circuit. The discrimination obtainable, that is, the closeness to which frequencies may be read off, depends upon the scale of the variable condenser, the values of inductance and capacity used, and on the decrement of the circuit (including the effect of the detector).

A wavemeter for use at a transmitting station in which these points have been carefully studied is shown in Fig. 221. The wave to be measured is picked up by *A*, which is coupled by minute condensers *D*, *D*, to the resonant circuit. This consists of a variable condenser *E* in parallel with a fixed condenser (whereby an open frequency scale is obtained and mechanical changes in the variable minimised) and a series of plug-in coils. The whole construction is very solid, and mycalex insulation is used, this being very permanent mechanically and also a low-loss insulator at high frequencies.

Resonance is detected by the very loosely coupled valve

detector. Connected in this way with positive on the grid, and a small negative bias on the anode, the anode current rises sharply for small voltages impressed.

The very small condensers *G*, *G*, and the tapping switch *H* enable a close reading to be obtained in the following way. The main variable condenser is varied very slowly, tapping *H* frequently, until a position is found where tapping *H* produces no change in the galvanometer reading. It will be seen that the variable condenser plus *G*, *G*, is just as far below tune as the variable condenser alone is above tune, the condensers *G*, *G*, just "bridging" the resonance curve. In this way the steep sides of the resonance curve are used rather than the flatter

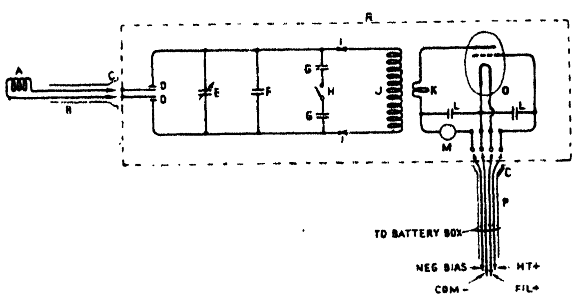


FIGURE 221.

top, in fact the tapping switch does semi-automatically what most observers usually do when finding resonance accurately.

Quartz Crystal Wavemeter. An accurate and convenient wavemeter for checking the frequency of a station or a number of stations all supposed to work on the same frequency can be made, using the arrangement of Fig. 222. The resonant circuit should have approximately the same natural frequency as the crystal, but need not be very low damped, as the accuracy of reading does not depend on this. When the transmitter frequency is raised the neon tube will commence to glow at the point *A* (Fig. 222), but will suddenly go out at *B* for a very small change *BC* of transmitter frequency, and will then continue to glow until *D* is reached. The "crevasse" at *BC* is due to the crystal oscillating with comparatively large amplitude at its natural frequency and

absorbing the necessary energy to do this from the resonant circuit with a consequent lowering of the voltage across it.

A Crystal Monitor. A useful type of monitor wavemeter for transmitters has been made by the Marconi Company. Briefly, the wavemeter consists of a receiver to which is coupled a crystal oscillator one of whose harmonics is set 300 cycles per second away from the frequency of the station it is desired to check.

The beat input between crystal harmonics and incoming signal energises an ordinary reed frequency-meter having a

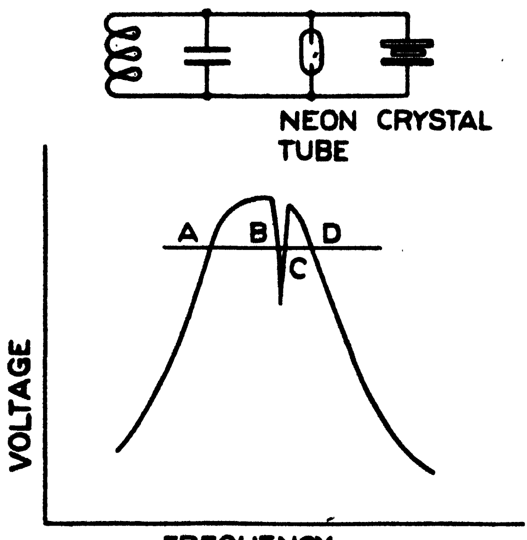


FIGURE 222.

set of reeds covering a frequency band of 300 $\pm$ 42 cycles per second with a central frequency of 300 cycles per second. Thus a visual check is obtained of the station frequency by observation of the frequency meter, and it is possible to observe frequency changes of as little as 2 cycles per second, by comparison of amplitudes of adjacent reeds; the sense of departure of station is also shown.

Frequency Checking Equipment. The above may be called reference methods, but where very accurate results are required it is safer to use an absolute method. Since frequency

is but the reciprocal of time, all absolute methods are based on a direct reference to time, and hence the essential feature of the measuring system is a very good clock. Some such absolute method is in any case essential for the calibration of wavemeters, but it is now becoming usual to check the frequency of transmitting stations directly, instead of using a wavemeter, as more accurate results can be obtained. Their transmissions are received at a frequency-checking station and then the frequency measured by an absolute method, which should be reasonably rapid and accurate to at least 1 in 10^{6}. Most administrations controlling wireless, and many official organisations, have now set up such checking stations.

The British P.O. Frequency Checking Station. The licensing and control of wireless stations in Great Britain is vested in the Post Office and in consequence this department maintains equipment to check transmitter frequencies. The primary frequency-standard is held at the Radio Laboratories, Dollis Hill, and a secondary standard is maintained at the Frequency Measuring Station situated at Colney Heath, St. Albans. The purpose of this station is to ensure that all the radio transmitters within the jurisdiction of the administration maintain their respective allocated frequencies within the prescribed tolerances and to act as a check where a case of interference occurs by or with a station in the administration. A brief description of the apparatus and methods employed at Colney Heath will be given here, but the reader should refer to published descriptions for detailed information.

The secondary frequency standard consists of a valve-maintained tuning fork, nominal frequency 1 kc/s, whose frequency is compared daily—by means of a land line connection—with that of the primary standard, the absolute frequency of the latter being determined daily in terms of mean solar time.

Frequency multiplying apparatus is associated with the fork to produce harmonic frequencies of the fork in the range 10 to 24,000 kc/s, and the principle of measurement is one whereby the frequency of an oscillator (termed the comparison oscillator) (CO) is set by adjustment of a condenser (IC) to the signal frequency and to two known harmonics of the fork respectively above and below the signal frequency, the latter

being determined by interpolation. The interpolating condenser (IC) is of special design in which backlash is avoided and in which the change of capacitance is directly proportional to the displacement.

The measuring apparatus is divided into two sets (see Fig. 223), the one for measurements in the range 10–1,340 kc/s, and the other for measurements in the range 1,200–24,000 kc/s. The long wave set is similar in principle to the short wave set which is outlined here.

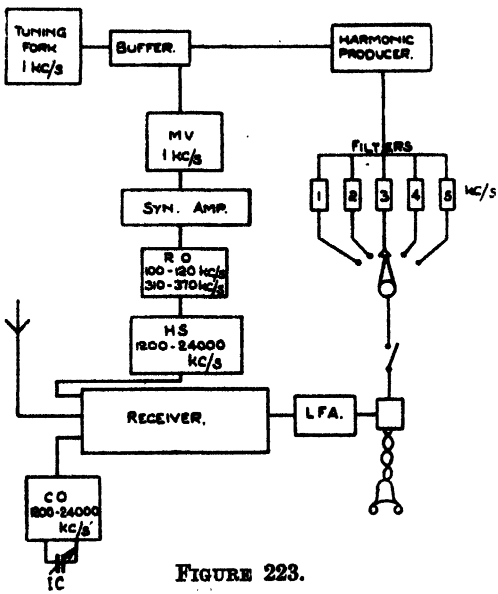


FIGURE 223.

MV	Mulivibrator.
RO	Reference Oscillator.
HS	Harmonic Selector.
CO	Comparison Oscillator.
IC	Interpolating Condenser.

For short wave measurements an output from the buffer stage is caused to control rigidly the fundamental frequency of 1 kc multivibrator. The wave form of the multivibrator is such that it is rich in harmonics, any one of which in the ranges 100–120 kc/s and 310 to 370 kc/s may be selected and amplified in the synchronising amplifier. The output of the synchronising amplifier is fed to the reference oscillator (ranges

100–120 kc/s and 310 to 370 kc/s), and this output is sufficient to hold the oscillator rigidly in synchronism with any selected harmonic provided the oscillator is adjusted approximately to the frequency of the harmonic. Harmonics of the reference oscillator are then produced, selected and employed as the reference frequencies for short wave measurements.

The procedure is best followed by an example and is more simply explained if the station frequency is assumed to be known. Let this be 17,270·5 kc/s. If the 339th harmonic of *MV* is selected and *RO* synchronised with it and the 51st harmonic of *RO* be selected, then the final frequency will be 17,289 kc/s (assuming the fork frequency to be exactly 1 kc/s). *CO* is now tuned to this and the reading of *IC* noted. The 50th harmonic of *RO* will give a final frequency of 17,238 kc/s and *CO* can be tuned to it and another point on *IC* determined. It will be seen that if *CO* is now tuned to the signal, the reading of *IC* will be between those previously obtained and hence the signal frequency can be determined by interpolation. The two known points on *IC* will be 50 kc/s apart, however, and this is too large for a precision measurement.

Greater accuracy is obtained in the following way. *RO* is set to the 314th harmonic of *MV* and its 55th harmonic selected giving a frequency of 17,270 kc/s. A 1 kc tone direct from the fork is now supplied to the telephones and by the "double-beat" method *CO* can be adjusted to 17,271 kc/s. Hence the signal frequency can be found by an interpolation on *IC* over 1 kc only. It will always be possible to get *RO* within ± 10 kc/s of the signal frequency and hence if fork harmonics up to the 5th can be applied to the telephones, it will always be possible to "step" up or down to the signal frequency. This method gives the unknown frequency in terms of the fork frequency to within ± 3 parts in 10^6 and the absolute accuracy of measurement is within ± 10 parts in 10^6.

Due to the fact that *MV* imposes a 1 kc/s modulation on all the harmonics selected from it, a further series of double beats can be obtained and the interpolation reduced to a few hundred cycles, but this refinement is seldom necessary.

Later frequency checking equipment [11] uses a 1 Mc/s quartz crystal in a bridge-stabilised circuit. It is then necessary to use a dividing circuit in order to compare with time signals.

Constant Temperature Chambers. Since these form an essential feature of some constant-frequency generators, a brief description of the methods employed for the maintenance of constant temperature seems desirable.

The chamber which houses the apparatus should be well lagged and completely enclosed, and in order to maintain the inside at constant temperature it is essential to increase the heat well above the average room temperature. Initially, fairly large boxes were used, the air of which was kept circulated by the use of stirring fans, but it is now found to be both more economical and more efficient to build two small lagged boxes, one within the other, the inner one being only just of sufficient size to contain the master oscillator unit. The inner box is then surrounded by a heater element, the mat type of heater being a very convenient form as it can be wrapped round the box. The control of the heater can be carried out by any convenient form of thermostat.

Simple types of thermostats consist either of mercury thermometers with sealed-in electrodes, or bi-metallic strip controls. The former are satisfactory if great care is taken in manufacture to ensure chemically pure mercury, perfectly clean glass, inactive electrodes and seals such as platinum, and if evacuated, inert gas should be sealed in the tube above the mercury.

Thermostat systems made in this way are extremely reliable and can control temperature to a small fraction of a degree centigrade.

Bi-metallic strip thermostats, which are in general not so accurate, are found in many forms, the usual form of which consists of a strip of bi-metallic substance which bends when heated due to the unequal expansion of the two materials in contact, this bending actuating a relay in the heater circuit and so cutting off the heat.

One good type of regulator can be made in the following manner. A steel base is made up fitted with steel blocks at each end. To these blocks is rigidly fixed a strip of metal of dissimilar temperature-coefficient to that of the steel base, and this strip is "set" in a definite manner. At the centre of the strip is a platinum-iridium contact, and above the strip is mounted on an insulating bridge a contact screw with a

platinum-iridium point. In order that the contacts shall not carry heavy currents, they only operate a relay which controls the main heater lamps.

SELECTED REFERENCES

(1) *The Properties of Silica.* Sosman. Chemical Catalogue Co., New York. 1927.

(2) *Notation for Piezo-Electric Quartz.* The N.P.L. H.M. Stationery Office.

(3) *The Valve Oscillator.* Thomas. Chapman & Hall.

(4) *Improvements in Quartz Crystal Circuit Elements.* Lack, Williard, Fait, *B.S.T.J.* July, 1934.

(5) *Note on Piezo-Electric Crystal.* Koga. *P.I.R.E.*, Vol. 14; March, 1936.

(6) *Resonant Lines in Radio Circuits.* Terman. *Electrical Engineering*, Vol. 53; July, 1934.

(7) *Frequency Control by Lines.* Hansell and Carter. *P.I.R.E.* Vol. 24; April, 1936.

(8) *Generation and Utilisation of U.S.W.s.* Kolster. *P.I.R.E.*, Vol. 16; April, 1938.

(9) *Ultra-Short Wave Oscillators.* Black. *Electrical Comm.* April, 1939.

(10) *Triode Oscillators for U.S.W.s.* Gavin. *W.E.*, Vol. 16; 1939.

(11) *The Application and use of Quartz Crystals in Telecommunications.* Booth. *J.I.E.E.*, Vol. 38; June, 1941.

(12) *Bridge-Stabilised Oscillator.* Meacham. *B.S.T.J.* 1938. Page 574.

CHAPTER XII

ELECTRON OSCILLATORS

THE consequences of the fact that electrons take a finite time to travel between cathode and anode of a valve have already been discussed in Chapter X and will come up again in connection with receivers in Chapter XIV.

In the present chapter we discuss types of oscillation which may have a much higher frequency than can be produced by reaction circuits.

We have included in this chapter, as a matter of convenience, the dynatron oscillation of the magnetron although this is not dependent upon electron transit time and is not therefore an oscillation of the electron type.

Many investigators have worked upon electron oscillators, and the literature is very extensive. Due to the high frequencies generated, measurements are very limited and approximate and the interpretation of the complicated experimental data correspondingly difficult. As a consequence, conflicting theories have been put forward from time to time and a complete quantitative analysis is still lacking.

Electron oscillators have not, up to the present, been used to any extent in actual communication circuits but it is probable that increasing use will be made of them in the future, when the lower ultra-high frequency bands (which are now covered by circuits of more conventional design but employing specially designed valves) have been fully exploited.

Barkhausen-Kurz Oscillator. In 1919, Barkhausen and Kurz, whilst testing the vacuum of transmitting valves, placed a positive voltage on the grid and a small negative voltage on the anode. They found that a current was recorded in the anode circuit and, by attaching a Lecher wire circuit between grid and anode, ascertained that a short-wave oscillation was taking place. The frequency was mainly

controlled by the grid voltage and the oscillations were not dependent upon reaction in the external circuits.

We can best explain the Barkhausen-Kurz oscillation by considering a triode valve having a plane parallel-electrode system preferably such that the distance between cathode and grid is approximately the same as between grid and anode, the grid being of rather open mesh.

Assume a positive potential $+E$ applied to the grid and let the anode be at zero potential. This means we have similar potential gradients rising from both cathode and anode to the central grid. Since the cathode is the emitter, we shall obtain a D.C. grid current I_g, and the heat appearing at the grid is due to the energy given up by the kinetic energy of the electron stream, and is a measure of the work done on the electrons by the accelerating field between cathode and grid. Generally speaking many of the electrons will stream direct from cathode to grid, but in certain valves it is found that quite a large percentage of them will be found to execute an oscillatory movement within the valve before capture. This is because of the probability of an electron missing the grid in its passage; the consequent retarding field in the grid-anode space will cause the electron to come to rest at a point just short of the anode, from whence it will be accelerated back towards the grid once more. We can imagine certain electrons therefore executing a "to and fro" shuttle movement before final capture. The maximum velocity such electrons achieve and the frequency of oscillation will depend on the value of E, and the distance from cathode to anode primarily, whereas the average number of oscillations made will probably depend mostly upon the grid mesh. If we assume the electron just swings between cathode and anode we can get an approximate idea of the frequency of this oscillation as follows:

If it is assumed that the electron leaves the cathode at zero velocity and if v be the velocity of the electron in cm. per sec. when it shoots through the grid,

e = charge on electron in coulombs ($1{\cdot}59 \times 10^{-19}$).
m = mass of electron in gm. ($9{\cdot}04 \times 10^{-28}$).
E = grid potential in volts.
l_g = distance from cathode to grid.
l_a = distance from cathode to anode.

The work which has been done on the electron when it reaches the grid is Ee joules or $Ee \times 10^7$ ergs and this must equal the kinetic energy $\frac{1}{2}m\,v^2$ ergs.

$$v^2 = 2\,Ee/m \times 10^7$$

i.e. the maximum velocity, $v = 5{\cdot}93 \times 10^7 \sqrt{E}$ cm. per sec.

Assuming a linear field—the average velocity

$$\frac{v}{2} = 2{\cdot}97 \times 10^7 \sqrt{E} \text{ cm. per sec.}$$

Hence the time taken to travel from cathode to grid is $\dfrac{l_g}{2{\cdot}97 \times 10^7\sqrt{E}}$ and from grid to anode is $\dfrac{l_a - l_g}{2{\cdot}97 \times 10^7\sqrt{E}}$

Thus the time taken for a complete oscillation from cathode to anode and back will be

$$= \frac{2\,l_a}{2{\cdot}97 \times 10^7\sqrt{E}}$$

and the frequency f_e of oscillation is $14{\cdot}9\,\dfrac{\sqrt{E}}{l_a}$ Mc/s.

The corresponding wavelength will be

$$\lambda = \frac{3 \times 10^{10}\,l_a}{14{\cdot}9 \times 10^6\sqrt{E}} = 2010\,\frac{l_a}{\sqrt{E}} \text{ cm.}$$

The valves usually used for electron oscillators have a single, straight filament down the centre of a cylindrical, co-axial grid and anode. It is evidently mainly a geometrical problem to extend the simple theory previously given to meet this case, but the above formulæ are sufficient to enable an estimate of the oscillation frequency to be made.

The above relationships have been deduced for a single electron but it is evident that if detectable oscillations are being produced there must be a vast number of electrons carrying out this oscillating motion, in other words, there must be an oscillating space-charge between the electrodes. This distorts the potential distribution so that the equivalent position of cathode and anode are closer to the grid and in consequence the wavelength found experimentally is usually less than that given by the above expression.

Since the electrons are leaving the cathode in random fashion, it would appear that the net effect of the oscillations would cancel out, since as many are passing in one direction at any one time as the other, but it is found possible to co-ordinate their motion so as to extract a small percentage of the oscillating energy they possess and which they have derived from a D.C. supply.

Maintenance of Oscillations. The way in which the oscillatory motion of electrons can maintain oscillations in an external circuit is still the subject of controversy and we give below alternative explanations which have been advanced. We might mention, however, that the *B-K* oscillation is the result of a "velocity modulation" of the electron stream (see page 387) as distinct from the current-density type of change created at lower frequencies. Such velocity-modulated systems are now being intensively studied by a number of workers for ultra-high frequency purposes and when their action is fully understood it will lead to a better appreciation of the *B-K* oscillator.

For simplicity we assume a parallel-plane electrode arrangement, with the grid halfway between cathode and anode. Connect a circuit, tuned to $\frac{\omega}{2\pi}$ c/s, between grid and anode and suppose a small alternating E.M.F., $E \sin \omega t$ to be applied in this circuit.

For the first explanation, the value of $\frac{2\pi}{\omega}$ is taken as equal to the time for a complete oscillation of an electron, that is, from cathode through grid to anode and back to the cathode.

In Fig. 224 we trace the course of an electron, correlating the instantaneous values of $e = E \sin \omega t$ with electron position. On the left hand side is shown diagrammatically the plane, parallel electrode-system, the different positions of an electron against time being plotted downwards, and on the right, grid potential changes are shown correlated.

Consider first an electron which leaves the filament at the instant that e is a maximum negative. This electron will be accelerated to a less extent than normal and this means that energy is being given up to the circuit providing the A.C. voltage. After electron has passed through the grid, e has reversed

and the electron is therefore retarded more than normally and will not reach the anode, but it is still giving up energy to the external circuit. On the return journey to the gird *e* is still positive, and so the electron is accelerated more than normally and abstracts energy from the circuit. If we suppose, however, that this electron is now captured by the grid, then it will be seen that the oscillation of this electron has had the net effect of giving up energy to the circuit, the shaded portions showing the portion of cycle when this electron gives up energy.

In the same way the flight of electrons which leave the cathode at other instants of time may be studied and in Fig. 224, right, we trace an electron path which absorbs energy from

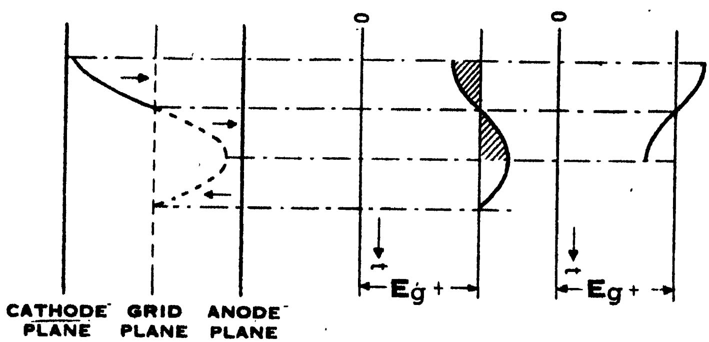


FIGURE 224.

the circuit and arrives at the anode, thereby contributing to the anode current. This current only flows when oscillations are occurring and may be used as a measure of their amplitude but it will be seen that the anode current produces damping of the oscillations.

The above shows the possibility of the electron oscillation supplying power to the external circuit, so that if the power thus supplied is sufficient (that is, greater than the circuit losses) oscillations may be sustained without an applied A.C.

Such an explanation fits in with the theoretical formula for the frequency previously derived, but it will be seen that the only electrons which yield up appreciable energy are those which leave the cathode during a small portion of the cycle and are captured by the grid on their return journey from the anode.

For the alternative explanation we assume that $\frac{2\pi}{\omega}$ is equal to half the time for a complete oscillation. The effect of $E \sin \omega t$ would average out on the D.C. stream of electrons travelling direct to the grid, but it would have a selective action on the random oscillations of those electrons executing a "to and fro" movement, depending upon the phase of the applied E.M.F. and that of the electron group considered, and we will consider the two limiting cases; first where the E.M.F.

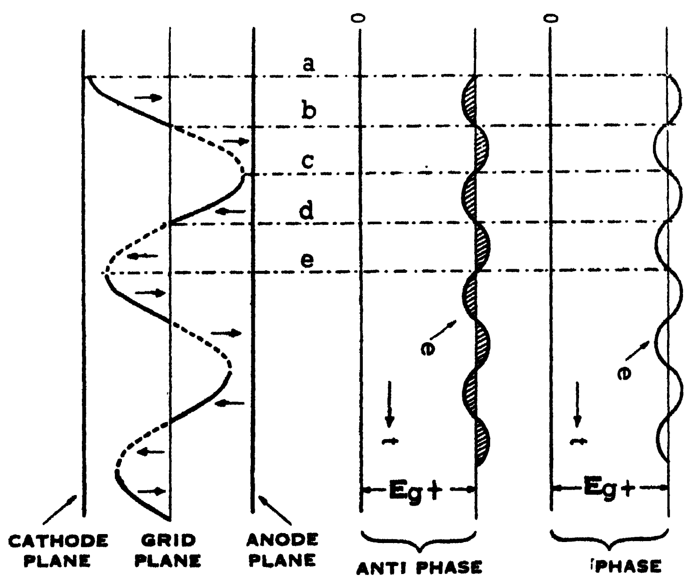


FIGURE 225.

is in phase opposition to the forward movement from the cathode; secondly where it is in-phase.

These two cases are illustrated in Fig. 225. On the left is plotted the path of an electron under the influence of the D.C. field alone against time (vertical) and opposite is the superimposed field of double frequency for the two phase conditions mentioned. Where the electron is in an accelerating field is indicated by a full line and where it is in a retarding field by a dotted line.

Consider the anti-phase case. It is clear from an examination of the figure that the superimposed E.M.F. is at all times

retarding the motion of the electron until its final capture. Thus from a to b, the acceleration is reduced because of reduction of E, from b to c the retardation will be increased because of the increase of E from c to d the acceleration back towards the grid is reduced because of the reduction of E, and from d to e the retardation is increased because of the increase of E.

The loss of motion is a measure of the energy delivered up to the LC circuit by the electron at its correct frequency. From this figure it will be clear why the superimposed frequency needs to be twice that of the electron excursion-frequency. Although from a geometric point of view a cycle will be from cathode to anode and back to cathode, from an A.C. point of view this represents two cycles of events, because in this time the electron has been twice accelerated and twice retarded.

Consider now the in-phase case. The phase of the superimposed E.M.F. is now seen to be such that it accelerates the travel of the electron at all times and this means it will swing to a greater amplitude and in all probability be carried over to the anode on its first outward journey, the group of electrons in this category thus showing their presence by anode current. The increasing acceleration given of course means energy is extracted from LC, but since such electrons are removed from the circuit right away, there is a net balance (in respect of the two groups mentioned) of energy being delivered to the external circuit, considering an equal number of electrons to exist in each group. Electron groups having intermediate phase conditions may be considered as falling into one or other of the cases mentioned.

The general effect is then that a number of electrons are eliminated from the circuit quickly in the form of anode loss and current, and a number surge to and fro within the valve, delivering up a proportion of their energy to the LC circuit each cycle and it may be imagined that this surging group tends to draw into synchronism electron groups which are not too far removed from it in phase, so that the initial random phase no longer exists.

Oscillations maintained in this way would evidently have a frequency double that given by the equation of page 362, but on the other hand, such a type of oscillation would appear to

have a greater effect on the external circuit than the type suggested in the first explanation.

It has been found experimentally that oscillations can be produced at about twice the frequency corresponding to a complete electron oscillation, as well as at that frequency.

Either explanations show the possibility of an electron oscillation yielding up energy to an external circuit, but the conditions in the practical cases are somewhat different. Thus cylindrical electrodes will normally be employed and for the best results a ratio da/dg of 3 gives the best results.

In order to obtain good results from a B-K. oscillator it is necessary to employ circuits of very high Q value, such as a resonant line. Although the circuit can be connected between grid and cathode, it is more usual to connect it between grid and anode, as from the H.F. point of view the anode is at the same potential as the cathode. Such a circuit is shown in Fig. 226, the wavelength being adjusted by the position of the slider.

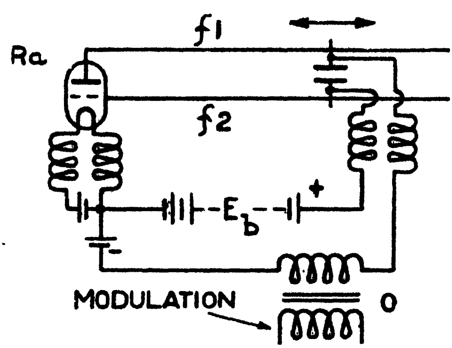


FIGURE 226.

Valves used for this purpose are usually made with a tungsten bright-emitting filament, as this is the most satisfactory type of filament to stand up to heavy electronic bombardment. Even with a tungsten filament this bombardment causes valve life to be short. It is also found that the cathode temperature is critical and this may have something to do with the control of the number of oscillations the useful electrons can make before capture.

Some of the earlier workers on this subject considered that two quite distinct oscillations were possible, one of which was entirely independent of the external circuit and another type in which the frequency was mainly dependent upon this external circuit. Later work, however, has shown that this distinction does not exist but that for certain adjustments the external circuit has rather more effect than for others.

During the latter part of 1931, considerable original work was done by Marconi and G. A. Mathieu on waves below 50 cm, using a modified form of B-K. oscillator, which is designed so that, if desired, a number of them can be built to form an array system and thus produce a narrow beam, Mathieu's oscillator, which is a two-valve type, is shown in Fig. 227 and with this circuit no difficulties have been experienced in getting ample power into the aerial system efficiently.

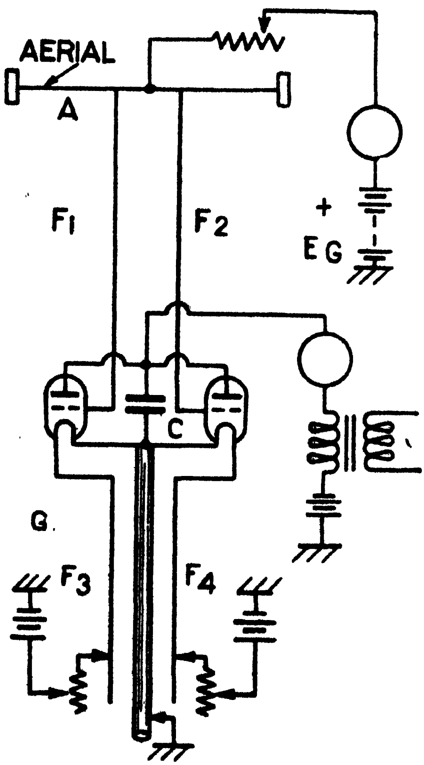


FIGURE 227.

It will be observed that the output circuit is connected between the two grids, and the two filaments are joined by short leads, as are the two anodes. Anodes and filaments are all held at the same H.F. potential by the condenser C.

The other filament leads are prolonged to Lecher wires F_3 and F_4, the length of these wires being so adjusted so that the filaments can be supplied with heating current at the voltage nodal-points, and for the proper working of these oscillators this adjustment is very important. The high-tension to the grids is fed through the centre point of the half-wave aerial and, to make a practical design of apparatus, the common filament-anode lead is made of a copper tube, which serves to support the whole transmitter and to screen the anode and grid supply leads, which are run through the centre of the tube. The special valves used in this oscillator have a life of several hundred hours when producing a frequency of 1000 Mc/s. The modulation voltage is introduced

into the anode circuit by means of a transformer which must have a low resistance.

The Resonant-Grid Oscillator.[2,3] A somewhat different type of oscillator, Fig. 228, is used by the I.T. and T. Corp. on the circuit established by them across the Straits of Dover and working on 17 cms, the shortest wavelength in regular use.

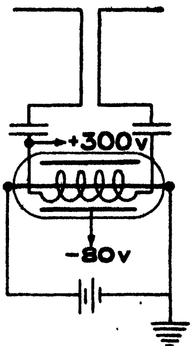


Figure 228.

The construction of the valve is similar to that of a valve suitable for a B-K. oscillator except that the grid has a connection at either end and must have no supports running along it, such as are usual in many valves. A Lecher-wire type of output circuit is connected between the two ends of the grid instead of between grid and anode.

Oscillations are found to occur when suitable potentials are applied, the condition always being such that all the electrons emitted by the filament are crossing over the filament-grid space.

With this type of oscillation the important valve dimensions which determine the frequency are those of the grid, that is

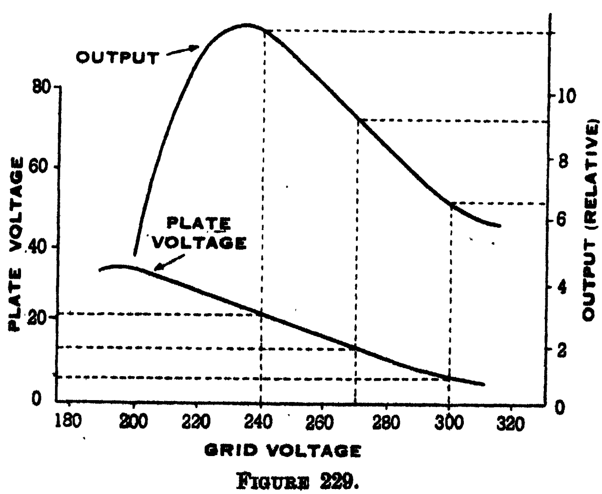


Figure 229.

length of wire, number of turns, etc. It is found that the oscillating frequency can be kept the same if grid and plate

voltages are both varied but the output will be different for each pair of values chosen. Typical curves are shown in Fig. 229 and it will be seen that if a modulating voltage is applied in the right proportions to both grid and plate, amplitude modulation will occur. For the characteristics shown, if the grid and plate supply voltages are +270 and −14 respectively and peak voltages at the modulation frequency of 30 and 8 volts respectively are applied to the grid and plate, we shall obtain a modulation of approximately 30%. This will be the maximum depth of linear modulation obtainable, but the amplitude modulation so obtained will be free from frequency modulation.

The production of oscillations of this type are explained by Clavier as being due to the grid acting as a transmission line having negative leakance (see page 369) and he has carried out an approximate analysis to show the possibility of this negative leakance arising, due to the transit time of electrons between filament and grid.

The Magnetron. As first produced by Hull in 1920 for long-wave amplification work, the magnetron valve consisted of a diode having a straight filament and cylindrical anode, the valve being placed inside a solenoid so that current through the latter produced a field co-axial with the filament. The term "Magnetron" is now employed however for a type of valve used for ultra-short wave work, and is usually made with the anode divided longitudinally into an even number of segments, and the field is generally applied by means of an electro-magnet or a permanent magnet.

Zacek, in 1924, was the first to show that such a valve could be used to produce very high-frequency oscillations, and since the introduction of the divided anode its behaviour has been widely studied by a number of workers, although full agreement is not yet reached as to its mode of operation in certain cases. With a divided-anode type of magnetron there are three types of oscillation which must be considered, called "Electronic," "Resonance," and "Dynatron."

Consider a diode valve with a cylindrical anode (not divided) and co-axial filament, subjected to an electrostatic field by positive potential on its anode, and an "in line" magnetic field produced either by a solenoid wrapped around the envelope,

or by placing the valve between the poles of a permanent magnet. If we keep the voltage on the anode constant and vary the magnetic field, we obtain a characteristic as shown by

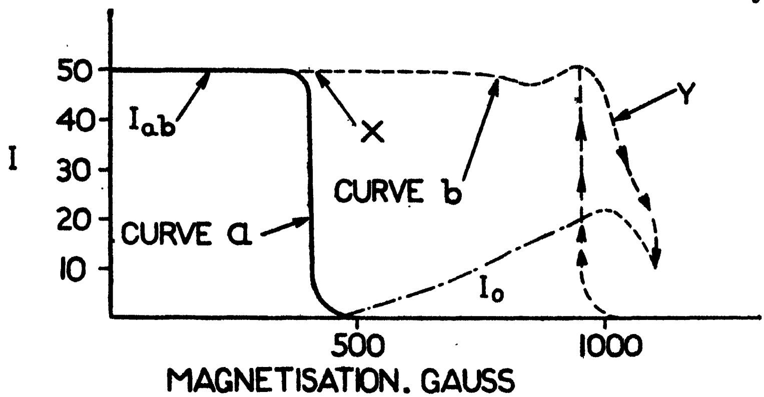
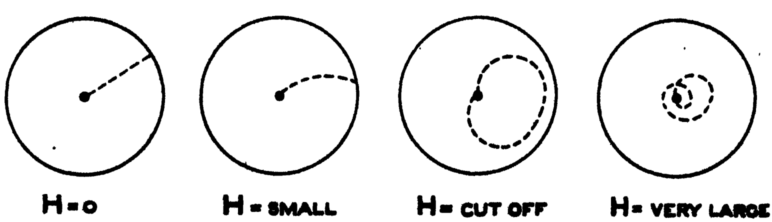


FIGURE 230.

curve *a*, Fig. 230. It will be observed from this that there is a critical value of field where the current falls abruptly to zero.

The shape of the characteristic obtained may easily be understood by studying the probable electron paths sketched in Fig. 231. When the field is weak, its only effect is to cause the electrons to follow curved paths between filament and anode but the number arriving at the anode is unaltered.



ELECTRON PATHS IN FULL ANODE MAGNETRON

FIGURE 231.

The force on an electron moving in a magnetic field is given by $H\,e\,v$, where H is the field strength, e is the charge on an electron and v is its velocity. The direction of the force is always normal to the direction of motion.

When the field is sufficiently increased, the electron paths are so curved that they become tangential to the anode, and the electrons just fail to reach it. Hence the anode current suddenly drops to zero. The sharpness of cut-off actually obtained depends upon the accuracy with which the electrodes are constructed and the uniformity of the magnetic field.

The random surging around of the electron stream between cathode and anode is a similar action to that experienced by the electrons in the positive grid triode, and the valve when set in such a condition has clearly the same possibilities as an electron oscillator.

(*a*) **Electron Oscillations.** Thus if a full cylindrical-anode magnetron is adjusted to "cut off" and a small superimposed alternating E.M.F. applied between anode and cathode; then the phase of this super-imposed E.M.F. with various groups of electrons may be considered as was done with the B-K. oscillator. If the phase is such that it retards the electron on its outward journey whilst it is in an accelerating field, it will still retard it on the return journey, because the E.M.F. reverses at the same time as the field changes from an accelerating to a retarding field, and this electron group will give energy to the circuit. Whereas if the phase of E.M.F. is such that it increases the acceleration from the cathode, although energy is given to the electron, this group is immediately swept out of the circuit and there is a nett gain of energy to the circuit.

Since the electron oscillation cycle consists only of one outward accelerating path and one return retarding path, the frequency of oscillation is the same as that of the superimposed E.M.F.

As has been mentioned, the usual type of magnetron now has its anode split longitudinally into one or more pairs of electrodes. Using a valve with a two-segment anode, for electronic oscillations, the tuned circuit will be connected between the segments. The maintaining action is fundamentally the same except that, owing to a small distortion of field near the slits, the two halves alternately impulse the circuit and thus the interval between impulses is halved, although the frequency of oscillation is the same and is determined principally by anode diameter and E_a.

The equation of motion of a single electron leaving the

filament of a magnetron adjusted to cut-off can be accurately determined and shows that the relation between anode potential (E), field strength (H) and anode diameter (cm.), at cut-off is

$$E = \frac{H^2 d^2}{181}$$

The time of transit of a single electron from filament to anode and back, under cut-off conditions can also be determined and hence the wavelength (cm.) of the oscillation found to be

$$\lambda = \frac{12{,}300}{H}$$

Or, by using the former equation

$$\lambda = 920 \frac{d}{\sqrt{E}}$$

Actually adjustments for cut-off and transit time will be affected by space charge conditions, and a complete solution is rendered difficult but it is found experimentally that

$\lambda H = 11{,}000$ is an approximate rule.

It is evident that the wavelength of these oscillations will be of the centimeter order and hence a line type of output circuit is most suitable, as shown in Fig. 232. The wavelength is, of course, determined mainly by the valve adjustments discussed above but varies to some extent with the circuit tuning, although the optimum output is obtained when the circuit tuning is adapted to suit the particular valve and valve adjustments. Fig. 233 shows a typical result with a magnetron having an anode of 1 cm. diameter. It will be observed that as the wavelength is reduced, it becomes necessary to increase anode voltage and magnetic field (as the equations show) and also filament current. It will be found that these adjustments are very critical, especially filament current.

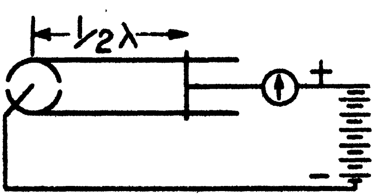


FIGURE 232.

In order to produce electron oscillations satisfactorily it is necessary to "tilt" the magnetic field, so that it makes a

small angle, usually about 6°, with the axis of the filament. Unless it is so tilted there will probably be no output or, if any is obtained it may not be of the "electronic" type. It is now assumed that the tilt is necessary to overcome excessive space-charge. Although space-charge limitation would not occur in an ordinary diode with the anode voltages being employed, it appears that in the magnetron the electron paths

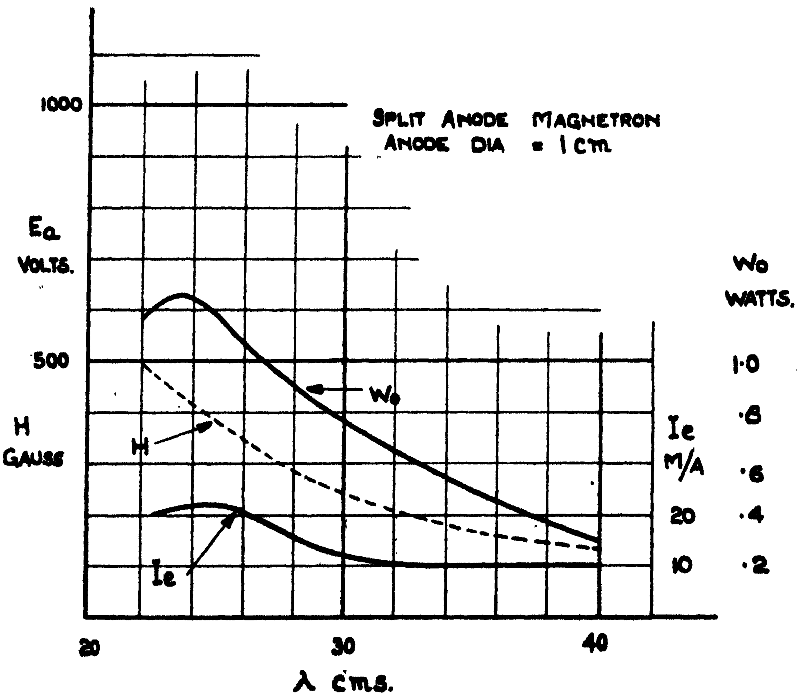


FIGURE 233.

lead to concentrations of electrons near the filament and if the field is tilted many of these electrons travel in spirals and are removed from the neighbourhood of the filament. An alternative to tilting the valve is to build into the valve two end plates or discs, one at each end of the cathode and set just outside the cathode and anode structure, leads from these two end plates being brought through separate seals. The application of a small D.C. potential has a similar effect to tilting the valve in the field, and in addition these end plates have been

used for modulation purposes by imposing a modulation voltage in series with the D.C. potential.

The efficiency of the magnetron when producing the electron type of oscillation is low—of the same order as the positive-grid oscillator but greater outputs appear to be possible and shorter wavelengths may be obtained. This is because the absence of a grid makes a very compact electrode system possible and there are no difficulties due to the grid overheating.

A frequency of 61,000 Mc/s (0·49 cm.) has been obtained by Richter, though the output was only about $2{\cdot}5 \times 10^{-7}$ watts for an input of 2·4 watts.

(*b*) **Resonance Oscillations.** The maximum frequency of this type of oscillation is about 1,200 Mc/s (25 cm.) and frequencies as low as 6 Mc/s (50 metres) may be obtained. This type of oscillation differs from the dynatron type to be described because the frequency is dependent upon E and H as well as on the tuning of the external circuit, and it differs from the electron type previously discussed because much lower frequencies can be obtained. There are reasons for supposing that the electron oscillation is merely a special case of the resonance type and that both depend essentially upon the transit time of electron movement within the valve.

For the lower-frequency resonance oscillations H is considerably above the cut-off value and it is then easy to distinguish between resonance and electron oscillations but (as will be seen from the next section) less easy to distinguish between resonance and dynatron oscillations which also occur with a high field value. At the highest frequencies, however, the electron and resonance oscillations merge together, both occurring near the cut-off point.

If a magnetron is arranged to produce resonance oscillations, then its anode current will vary with H as shown in Fig. 234. As seen previously, I is constant until the critical value of field is reached (A), when it begins to decrease rapidly. The electron oscillations previously discussed would occur in this region. At B the resonance type of oscillations commence and when these oscillations are occurring I rises again to a maximum at C, which is also the point of maximum output.

In the case of the full-anode magnetron, it is found that the wavelength (in cm.) of resonance oscillations is given by

$\lambda = K\frac{H}{E}$, where K is about 400 for a valve having an anode diameter of 1 cm. Shorter wave oscillations may also appear, such that $\lambda = \frac{K}{n} \cdot \frac{H}{E}$, where n is 1, 2, 3, etc., values up to 7 having been obtained.

If a magnetron has P pairs of segments, the most prominent oscillation is given by $\lambda = \frac{K}{P} \cdot \frac{H}{E}$, though the longer wavelength as given by the full-anode valve may also appear. Hence shorter wavelengths may be obtained by the use of a

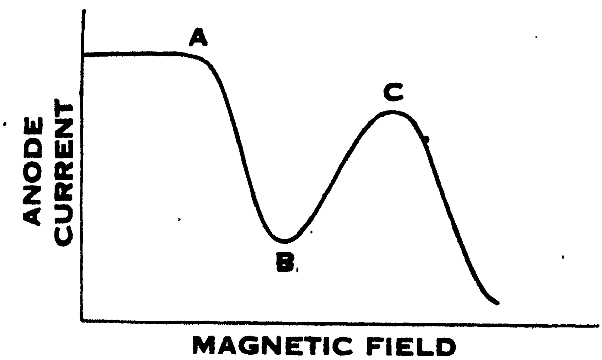


FIGURE 234.

number of segments, 2 or 4 being usual, but 8 and 12 have been employed.

The value of K is approximately constant for longer wavelengths but decreases considerably for the shortest wavelength obtainable. Thus McPetrie and Ford [14] found that $\lambda \frac{E}{H}$ for a magnetron having a 4-segment, 1 cm. anode, was very nearly constant at 200 down to 250 cm. but then fell rapidly, being 100 at 100 cm.

The same investigators find that in the 2-segment magnetron the so-called electronic and resonance oscillations merge at the shortest wavelengths. In order to determine with any definiteness the wavelength relationships outlined above, it was found necessary to limit the oscillations to a very small amplitude, by reducing the filament current.

The impedance between the segments of a 2-segment magnetron when it is producing resonance oscillations has been measured by Hervey[16] and found to consist of a negative resistance together with a reactance. The variation of negative resistance as shown in Fig. 235 gives a formal explanation for

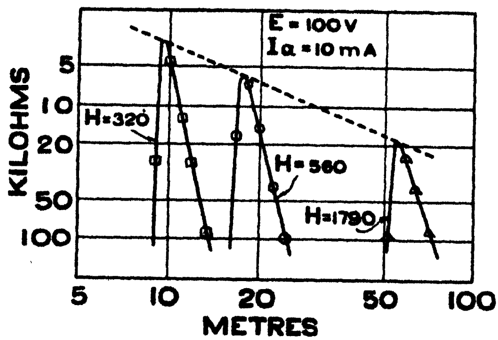


FIGURE 235.

the existence of resonance oscillations, but we have still to consider a physical explanation for their maintenance.

Since each frequency requires a different value of H and E and these control the electron paths, the mechanism of oscillation would appear to be dependent upon the transit time of electrons, that is the mechanism is similar to that of the electron oscillation but the frequencies generated can be, as we have seen, much lower than any possible transit time direct from filament to anode.

An explanation has been put forward by Gill and Britton[6] for a split anode magnetron which supposes that the electron paths are somewhat as shown in Fig. 236. The time taken for the electrons to progress in a number of orbits across the face of one half segment before capture by the other, can be seen to be comparatively long and if the alternating voltage is superimposed upon the steady potentials of the segments, then adjustment of magnetic field can so arrange the electron orbits that an electron takes half a cycle of the applied voltage to

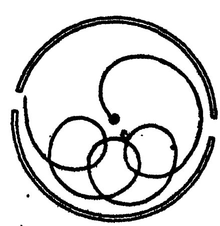


FIGURE 236.

travel from one gap to the next. As a result, at each gap—where the effect of the alternating field is strongest—the electron is urged nearer to the lower potential segment (in much the same way as illustrated by Fig. 238), thus contributing to the negative resistance effect

McPetrie and Ford have shown that resonance oscillations can be produced in a full-anode valve, and splitting up the anode into segments merely reinforces certain oscillations.

(*c*) **Dynatron Oscillations.** This type of oscillation is only possible with a split-anode type of valve, and is not an electron

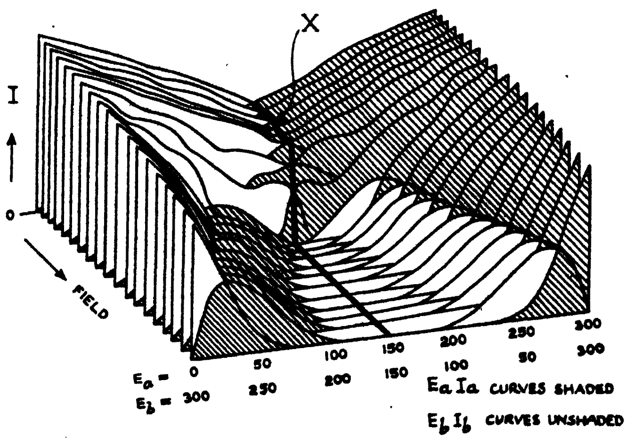


FIGURE 237.

oscillation at all. Consider a 2-segment magnetron valve. If we connect the two segments together and plot a characteristic of anode current-magnetisation, we obtain of course a curve similar to that shown in Fig. 230 (*a*). If, however, we separate the segments, and, for different field values, vary the voltages to each segment, we obtain a family of curves as shown in Fig. 237, from which it is seen that for the range of values of H beyond "cut-off," the segment at the lower voltage takes more current than the other. The reason for this can best be shown by considering Fig. 238 which shows the electric field when the segments are at different potentials. In Fig. 238, p gives the direction of force due to the electric field, f that due to the magnetic field, which is always at right angles to the direction of motion of the electron, and the full line sketches the

probable path of an electron which would have missed the anode and returned to the filament if both segments had been at the same potential E. It will be observed that the potential difference $2E_1$ between the segment distorts the field considerably near the gap and as a result the electric field has a component retarding the motion of the electron at this point in its path. This reduction of velocity decreases the magnetic force on the electron and reduces its curvature of path so that it arrives on the lower-potential segment.

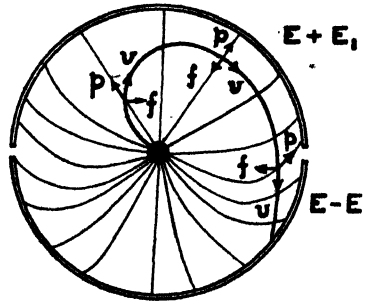


FIGURE 238.

Kilgore[9] has plotted the equipotential lines for a 2-segment magnetron when there is a considerable P.D. between the segments and shows that in this case the electron path may be a spiral, the electron arriving on the lower-potential segment, as in the case pictured above. By placing a small amount of argon in the envelope of the magnetron Kilgore was able to photograph the electron paths due to the ionisation produced.

We will now consider the action of such a valve assuming that between the segments we place an LC circuit whose

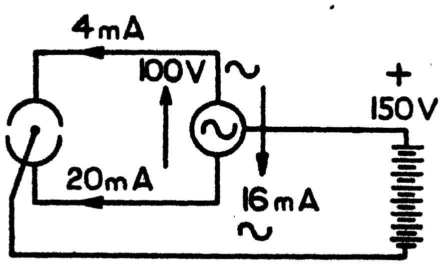


FIGURE 239.

frequency is low compared with the electron transit time. For the present we can therefore consider the circuit as representing an alternator of zero internal resistance, giving say a peak voltage of 100, as indicated in Fig. 239.

If it is assumed the field is set to beyond "cut-off" as shown by the most forward curve of the series of curves in Fig. 237, then at the peak condition of A.C., segment "*a*" at 50 volts is taking 20 mA, and segment "*b*" at 150 volts is taking 4 mA. Thus, although the alternator is producing a voltage of 100, it is being driven as a motor, since the current flow is in the opposite direction to the generated voltage.

It is now an easy step in the discussion to replace the alternator by a tuned circuit (Fig. 240) which, when it is oscillating will produce an alternating potential difference between the segments. If the "motoring" effect previously mentioned is sufficient, so that the magnetron gives energy to the circuit, then oscillations will be maintained, the energy coming from the anode D.C. supply.

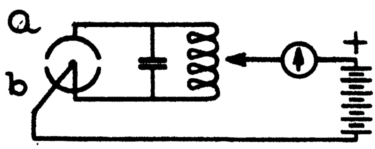


FIGURE 240.

Of course, another way of considering the production of these dynatron oscillations is to use the negative-resistance concept. Since the voltage between the segments is in the same direction as the current which appears to flow between the segments, instead of being opposed as in an ordinary resistance, we regard the impedance between the segments as being a negative resistance which can cancel the positive resistance of the circuit placed between the segments, and oscillations can thus occur.

This type of oscillation has evidently nothing to do with electron transit time, since it is due to a property which is apparent in the static characteristics, and the frequency is determined by the circuit connected between the segments. When these dynatron oscillations are taking place, a relation between total anode current and *H* will follow curve *b* of Fig. 230, cut-off occurring at a considerable higher value of *H*, this being particularly marked when the load is small.

It will be of interest to examine the factors affecting the

efficiency of a magnetron producing this dynatron oscillation. An examination of the static characteristics of Fig. 237 shows that for field values just beyond the "cut-off" the negative slopes commence comparatively near the equipotential point and quickly rise to a peak. As the field is increased, it requires a bigger difference of voltage before current passes through that segment having the lower voltage and the peaks of current are further removed from the equipotential point. This indicates that greater efficiency will be obtained with greater values of magnetisation because maximum current passes through the segment at a lower potential. This is found to be so in practice but, because no current passes until a considerable change of voltage is created, it is difficult to start oscillations with a high field but by a subsequent increase of field, larger efficiencies may be obtained. This non-reversibility, however, precludes efficient modulation of the anode potential.

The conversion efficiency of the dynatron oscillator is high, values of 50% being obtainable at 100 Mc/s and 35%–40% at 150 Mc/s, but it should be pointed out that the magnetic field is an essential part of the circuit, and as this usually requires appreciable power the overall efficiency is reduced by some 15%.

The high efficiency of the magnetron as a dynatron oscillator can only be maintained whilst the generated frequency is well away from the transit-time frequency. Kilgore has shown experimentally that if the transit time is less than 1/15th of the period, the efficiency falls to 30% and when it is 1/5th of the period the dynatron oscillation begins to fall. As the electron-oscillation frequency range of a given valve can be calculated with fair accuracy from dimensions and circuit constants, it is a simple matter to avoid overlapping of these two types of oscillation.

The adjustments of this oscillator are very simple as it is only necessary to set the circuit to the frequency required, adjust the magnetic field to an approximate value, and bring up the anode voltage until oscillations are obtained, which are indicated by a sudden rise in anode current : or conversely, the anode voltage may be set to the full value right away and the field varied, but always from a high value to a lower

one, until oscillations are obtained. Starting from a low field and increasing may overload the valve by causing excessive anode loss.

Filament Over-Heating in Magnetrons. It is frequently found that when a magnetron is producing either dynatron, electronic or resonance oscillations, the total anode current is greater than for zero magnetic field. This is because the filament is bombarded by returning electrons and its temperature is thereby increased and the emission becomes greater than would be produced by the heating current alone.

Since increased anode current may produce a further rise in filament temperature, and so on, the effect becomes cumulative and may result in the destruction of the valve. It may be necessary to insert a resistance in the anode circuit so that the current may be limited. In the case of the valve used for dynatron work, the effect can be prevented by placing two guard wires parallel to the filament, and in line with the anode slit, a small negative bias being added to these wires.' Guard wires, however, are not used in valves intended for electron oscillations as they prevent the oscillation from taking place. With the dynatron valve, overheating of the anode supports may also be serious, due to a similar cause, namely excessive bombardment, and this being prevented by placing on each support a small cylinder insulated from the support, which acquires a negative potential by accumulating electrons, and so prevents the main support from heating.

Frequency Stability of Magnetrons and Modulation. One of the chief disadvantages of the magnetron for communication purpose is its poor frequency stability under modulation conditions. The alteration of either anode voltage, field, endplate voltage, or guard wire voltage, for the purpose of modulation results in a large, unwanted frequency-modulation appearing, together with the amplitude modulation. A vast number of patents disclosing modulation circuits of one kind or another have been filed but these are in most cases clumsy or impracticable, and it is this inability to obtain successful modulation that appears to have retarded its progress as a practical device for producing ultra-high frequencies for communication purposes. It is, however, an efficient H.F. generator and in Chapter XVIII is shown a

circuit which has been developed for ultra-high frequency therapy work.

Considering for instance the dynatron type of oscillator, amplitude-modulation (accompanied by some frequency-modulation, as has been mentioned) may be carried out by varying the anode voltage, but we must avoid the discontinuity previously mentioned at high field values. Fig. 241 shows a set of curves connecting output current with E_a for different values of H, and it is seen from these curves that the curve $H = 500$ is linear over a fair range, although with this value

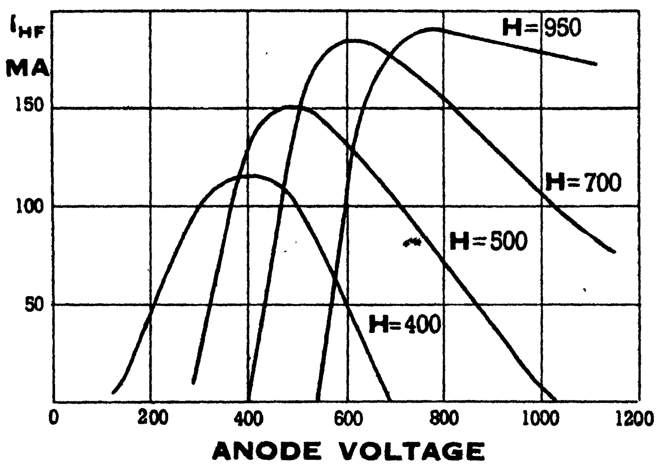


FIGURE 241.

of field maximum output cannot be obtained. With the maximum output curve the system will be discontinuous. Only shallow modulation at a reduced output is therefore possible.

This undesired frequency-modulation occurs, of course, when anode modulation of an ordinary triode self-oscillator is carried out and is due to the change of anode resistance with the changing anode voltage. Although the percentage frequency-change is very small, yet at ultra-high frequencies this represents a very large actual frequency change and can therefore seriously interfere with telephone modulation.

With the triode valve such a frequency change can be eliminated, or largely reduced, by the application of a modulating voltage in opposite phase and of appropriate amplitude,

on the grid of the valve, such as is done in the case of the resonant-grid oscillator circuit. Similarly, in the case of the magnetron valve, if the magnetic field can be varied at modulation frequency but in opposite phase to the modulation impressed on the anode, frequency modulation can be eliminated. Circuits have been suggested for carrying this out but they are difficult to apply economically.

Grid-Controlled Magnetrons. Magnetrons have been constructed with a grid-electrode for the purpose of aiding the modulation problem. Modulation may then be carried out either by varying the grid-voltage alone, or by applying modulation to both grid and anode in opposite phase. Fig. 242 shows the output-current-grid-voltage curve of a grid-controlled magnetron working on about 180 cm. wavelength, from which it is seen that deep, linear modulation is possible, but it is accompanied by frequency modulation.

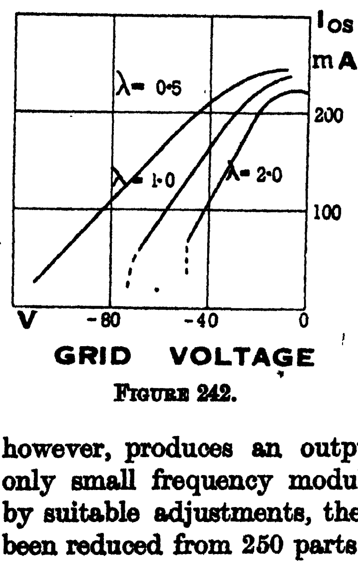


FIGURE 242.

The simultaneous modulation of grid and anode in anti-phase, however, produces an output linear over a range, but with only small frequency modulation. It has been stated that by suitable adjustments, the frequency modulation effect has been reduced from 250 parts in 10^5 to 50 parts at deep modulation, and if the modulation depth is reduced it is possible to reduce the frequency change to a value of 10 parts in 10^5. Such a figure is, however, not too satisfactory for telephony at ultra-high frequencies.

Special Types of Magnetron. The problems of the production of appreciable power at very high frequencies is much the same with the magnetron valve as with other oscillators—that is to say, to provide the emission we require a large cathode and to deal with the anode loss we require an anode system that will radiate heat efficiently, whilst these electrodes should be small if very high frequencies are desired. Since the magnetron seemed such a particularly efficient generator

of very high frequencies it is natural that a considerable number of types should have been evolved for producing appreciable power at very high frequencies.

Linder [12] has constructed a 2-segment valve in which the segments are one-quarter wavelength long and short circuited at one end and with this he produced a power of 15 watts at 8 cms.

Kilgore [9] attacked the problem of providing a comparatively large output efficiently at somewhat lower frequencies and has constructed a valve producing 100 watts at 600 Mc/s (50 cm.) with an efficiency of 25%. In this valve, as shown in Fig. 243, a massive Lecher-wire circuit of large thermal capacity and dissipating surface is placed within the valve envelope and

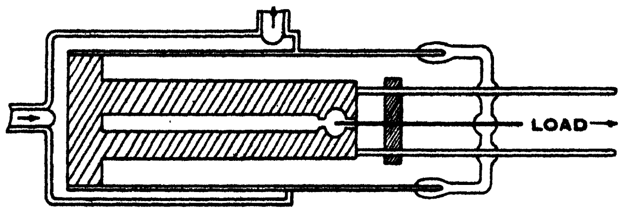


FIGURE 243.

water-cooled. By this means a large output is possible even though the size of actual anode is quite small.

Water-cooled valves have also been produced, in somewhat different fashion by Pfetscher and Puhlmann, and others; and a number of workers have suggested designs, in which the internal dimensions of the segments are a specific function of the wavelength, so as to facilitate the coupling of the internal energy with the external circuit.

Circuit arrangements are sometimes used so that part of the magnetising winding carries the anode current. One of the authors has built a dynatron type of oscillator in which practically the whole of the magnetisation is provided by the anode current, only a final setting being obtained by a subsidiary low-voltage winding. Such an arrangement prevents overloading of the valve as any rise of feed increases the field. It has also been found that with a valve used for resonance oscillation such a circuit is much more stable. With the ordinary circuit, if the magnetron is adjusted to produce

the maximum output and then the load is reduced, the anode quickly becomes overheated, but with the differential circuit this does not occur.

Electron-Deflection Oscillators. The extensive development in recent years of the cathode-ray oscillagraph and other devices in which an electron beam is deflected, has led to attempts to construct amplifiers and oscillators in which a beam of electrons is deflected by the input voltage instead of varying the number of electrons in a stream, as in the usual type of valve.

The essentials of such an arrangement are shown in Fig. 244 which may be a high-frequency amplifier if the deflecting plates are supplied from an external source, or a self-oscillator if their voltage is derived by coupling back from the output circuit. Evidently, the alternating voltage between the

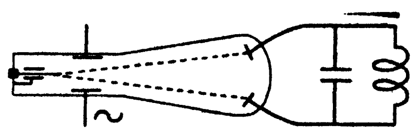


FIGURE 244.

deflecting plates will result in the electron beam "switching" from one anode to the other, thereby giving impulses to the resonant output-circuit.

Arrangements of this kind have suffered from the defect that, if the beam was made long enough so that the switching action took place with a low voltage between the deflecting plates, then the beam had a very high resistance and it was therefore difficult to obtain appreciable power and difficult to design an output circuit of sufficient resistance. In other words the mutual conductance of the arrangement compared unfavourably with that of the ordinary valve.

It has been shown that the input impedance of ordinary valves decreases greatly at high frequencies, and it can be shown that at such frequencies the impedance between the deflector plates may be much higher than the input impedance of an ordinary valve. The deflection-oscillator and amplifier may therefore find application at very high frequencies and attempts are being made to improve its mutual conductance.

Velocity-Modulated Beams. A second alternative to the conventional method of controlling an electron stream by varying the number of electrons, is to vary their velocity and such an electron stream is said to be "velocity-modulated." Control of velocity modulation can remain efficient, and the control electrode be of high impedance, even at extremely high frequencies, in contrast to the behaviour of the conventional control-grid. Fig. 245 represents an electron stream of uniform density and travelling with a uniform velocity between the electrodes *AB*, which have a steady potential on them equal to that which accelerated the electrons and produced the beam.

We will now consider the application of an alternating voltage e between A and B during a half-cycle of the voltage.

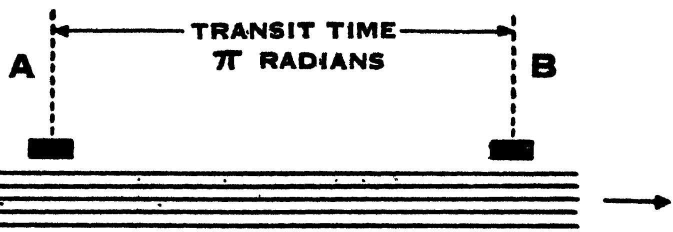


FIGURE 245.

A convenient way of expressing the time taken for the electrons to travel from A to B is to specify it as an "electrical angle" with respect to the frequencies applied to the electrodes, so that in this case we can say the transit time between A and B is π radians.

The value of e is assumed small compared with the voltage E which has produced the electron stream. Any changes of velocity brought about by e will therefore be a small fraction of the mean velocity and it is assumed that the transit time A to B is still π radians. Fig. 246 shows the relationship of voltage in A and B at different times. Thus an electron which passes A at time t_1, will pass under B at t_3 and since e is zero at both these instants, its velocity will be unchanged.

An electron which passes A at t_2 will be accelerated whilst passing A and will travel with increased velocity to B which it reaches at t_4 (approximately) and is therefore accelerated again and travels on at a uniform but higher velocity.

By a similar argument, an electron which passes under A at t_4 will be retarded under both A and B.

On the right of B, therefore, the electron stream contains electrons which have been accelerated due to twice the maximum value of e, electrons which have been retarded by an equal amount, electrons which are unchanged in velocity, and electrons which have velocities intermediate between the above particular cases.

If now we allow the electrons in such a velocity-modulated beam to "drift" along a tube after passing the control-electrode system, it is clear that the faster ones will catch up on the slower ones and what was formerly a continuous stream of electrons will become sorted into groups so that regions

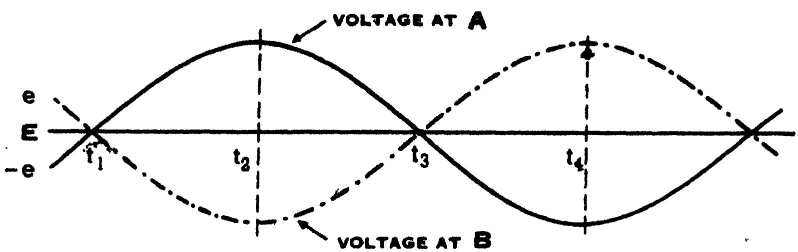


FIGURE 246.

where the electrons are closely packed and where they are rare will be travelling down the tube.

It will be appreciated that in order to achieve velocity-modulation with an electrode system of reasonable compactness it will be necessary for the velocity of the electron stream to be high, since the transit angle between A and B should be 180°. Quite a small alternating voltage on the control electrodes will produce a deep velocity-modulation provided the "drift" tube is long enough, since even the smallest difference of velocity that is produced at A and B will, if given time, effect a sorting process. This means that although only negligible power may be used to modulate the beam, it is possible to obtain large A.C. power from the beam by the adoption of a proper collecting device; so that such an arrangement can be used as an amplifier or oscillator.

Various methods for the collection of A.C. power from a

velocity-modulated beam have been devised, and tubes depending upon two different principles will now be described.

The "Klystron" Oscillator. A novel type of oscillator has been developed by R. H. and S. F. Varian[19] which shows promise as a means of generating considerable power at very high frequencies.

For good results at these frequencies it is necessary to use an unusual type of resonant circuit and this will first be discussed.[20] It has long been known that a metallic enclosure such as, for example, a copper sphere, can have an electromagnetic oscillation produced within it, in much the same way as a closed (or nearly closed) vessel can become an acoustic resonator and, more recently, the properties of such enclosures have been investigated and the name "rhumbatron" given to them.

If we suppose the walls of the enclosure to have zero resistance then there will be no field outside when an oscillation is taking place. Inside, there will be an alternating electric field and, in space and time-quadrature with it, an alternating magnetic field, and there will be currents on the inner surface of the enclosure.

It has been shown that such enclosures can have very large Q values. The definition of Q as L/CR hardly applies here but the equivalent definitions as given in the transmitting section, namely as kVA/kW holds.

For reasons which will presently appear, a suitable form of rhumbatron to employ in the klystron is that shown for the "buncher" and "catcher" so-called, in Fig. 247. In this case the electric field will be mainly concentrated across the neck of the rhumbatron.

Considering the arrangement of the klystron in Fig. 247, a stream of electrons is produced by an "electron gun" similar in arrangement to that in a cathode-ray tube, additional focussing of the beam being obtained by the outer web of the oscillating system, which as a whole is at the potential of the anode, and can be earthed. The few electrons which pass right through both buncher and catcher are collected on an electrode which may form an extension of the catcher, but this is not concerned in the production of oscillations. The seals and details of envelope to produce the necessary evacuated system are not shown.

Let us excite the buncher rhumbatron into oscillation and so arrange matters that the electrons pass right through the neck of it during a half cycle of the oscillation. This can be done by obtaining the correct value of D.C. potential. Then electrons which enter when the electric field due to the oscillation is opposing motion will be retarded whilst those which enter during the next half-cycle will be accelerated.

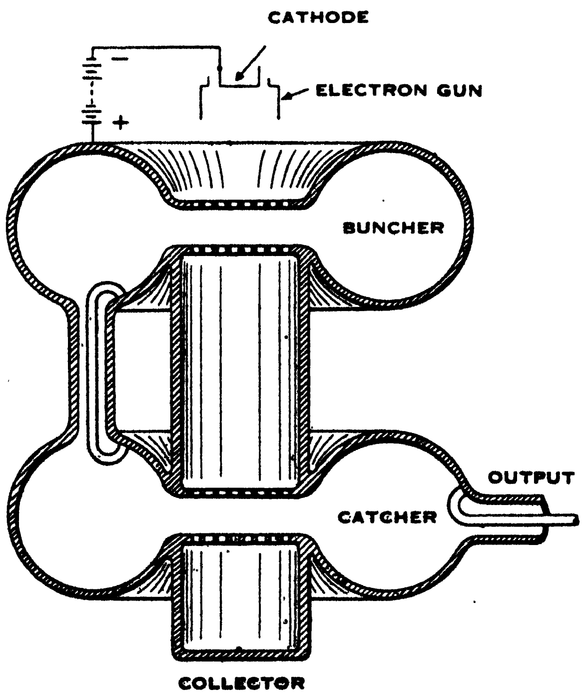


FIGURE 247.

The buncher therefore corresponds to the electrodes A B of our previous discussion and it is clear that velocity-modulation of the electron stream will be produced and the electrons will become bunched and rarefied as they drift down the tube beyond the buncher.

Let us now consider what happens at the second rhumbatron —the catcher. Ignore for the time being the concentric line connecting the rhumbatrons and suppose the catcher to be oscillating. If the phase of the oscillation is such that when

it is retarding the electrons, the bunched electrons are passing the catcher, whilst the rarefied electrons are passing it when its voltage is accelerating them, the electron beam will be yielding up energy to the catcher and a large conversion from D.C. to A.C. is thus possible. For high efficiency the catcher should have a high Q value so that a powerful electric field is produced across its neck and hence the electrons are retarded to nearly zero velocity in their passage across the catcher.

Evidently if the oscillation of the buncher is maintained by feeding back a small portion of the oscillating energy of the catcher by way of a concentric line of the right length to give the correct phase, we have a self-oscillating system.

It is necessary for the rhumbatron to be of the shape shown in order that electrons accelerated by convenient voltages may pass through it during a half-cycle of the oscillation. If a sphere had been used, for example (with the electron stream passing through its diameter), then its wavelength is only slightly greater than its diameter, so that, even if the electron stream had the velocity of light, it could not get through in a half-cycle of oscillation. With the shape of rhumbatron used, however, a voltage of about 3000 volts is sufficient to produce a sufficient velocity, about 1/10th that of light, for the correct phasing of the beam.

With the klystron oscillator outputs of 300 watts at 10 cm. (3,000 Mc/s) have been obtained and correspondingly larger outputs at longer wavelengths.

Velocity-Modulation and Retarding-Field Conversion. As an alternative to sorting out the electrons in a velocity-modulated beam by allowing them to drift along a uniform field, valves have been designed in which the sorting is accomplished by applying a retarding field to the velocity-modulated stream.[18]

If a low-potential electrode is placed in the path of the beam then it can be arranged that electrons travelling at the higher velocities will be collected by the electrode but those travelling at velocities below the average will be brought to rest before reaching the collector and will return back along the beam. Hence the effect of the retarding field has been to obtain from a velocity-modulated beam a returning electron stream which is varying in intensity at the operating frequency. This returning

stream can therefore supply A.C. power to any electrode upon which it impinges in the same way as the alternating anode-current in a conventional valve supplies A.C. power to a load.

The B-K. oscillator and the magnetron are really examples of a retarding-field type of velocity-modulated valve, but valves specially designed for the purpose, and incorporating a low-loss output circuit of the transmission-line type have been developed by Hahn and Metcalf and used as oscillators down to 5 cm., and as amplifiers and frequency changers at wavelengths as low as 37 cm.

SELECTED REFERENCES

(1) *Electronic Oscillators.* Megaw. *J.I.E.E.*, Vol. 72; April, 1933.

(2) *Micro-Ray Communication.* McPherson and Ullrich. *J.I.E.E.*, Vol. 78; June, 1936.

(3) *Production and Utilisation of Micro-Rays.* Clavier. *Electrical Comm.*, Vol. 12; July, 1933.

(4) *The Magnetron Short Wave Oscillator.* Megaw. *J.I.E.E.*, Vol. 72; April, 1933.

(5) *The End-Plate Magnetron.* Linder. *P.I.R.E.*, Vol. 24; April, 1936.

(6) *The Action of a Split-Anode Magnetron.* Gill and Britton. *J.I.E.E.*, Vol. 78; April, 1936.

(7) *Modulating the Magnetron Oscillator.* Groszowski and Ryzko. *P.I.R.E.*, Vol. 24; May, 1936.

(8) *Magnetron Valves for U.-S. Waves.* Megaw. *G.E.C. Journal*, May, 1936.

(9) *Magnetron Oscillators.* Kilgore. *P.I.R.E.*, Vol. 24; August, 1936.

(10) *Editorial.* G. W. O. Howe. *W.E.* July, 1936.

(11) *Electron-Beam Magnetrons.* Okabe. *P.I.R.E.*, Vol. 27; January, 1939.

(12) *Anode Tank-Circuit Magnetron.* Linder. *P.I.R.E.*, Vol. 27; November, 1939.

(13) *Resonance Oscillations in Magnetron Generators.* McPetrie. *J.I.E.E.*, Vol. 80; January, 1937.

(14) *Investigation of Resonance and Electronic Oscillations in Magnetrons*. McPetrie and Ford. *J.I.E.E.*, Vol. 86; March, 1940.
(15) *Magnetron Oscillator with Compound Field Winding*. Ford. *J.I.E.E.*, Vol. 86; March, 1940.
(16) *The Impedance of the Magnetron*. Hervey. *J.I.E.E.*, Vol. 86; March, 1940.
(17) *Deflection Methods for Amplification and Oscillation*. Colebrook. *W.E.*, Vol. 15; April, 1938.
(18) *Velocity-Modulated Tubes*. Hahn and Metcalf. *P.I.R.E.*, Vol. 27; February, 1939.
(19) *A High-Frequency Oscillator and Amplifier*. R. H. and S. F. Varian. *Journ. App. Phys.*, Vol. 10; May, 1939.
(20) *Resonators Suitable for Klystron Oscillators*. Hansen and Richtmyer. *Journ. App. Phys.* Vol. 10; March, 1939.
(21) *Theory of Klystron Oscillator*. Webster. *Journ. App. Phys.*, Vol. 10; December, 1939.

CHAPTER XIII

MODULATION CIRCUITS

When dealing with the subject of modulation theory, it was shown that the problem in its simplest form is to vary the amplitude of a high frequency carrier wave in such a manner that a line drawn through the tips of the waves forms an envelope shape which corresponds to the modulating signal. Further, to create this variation of amplitude necessitates work being done of amount which can most easily be determined by considering the side band analysis.

In the ideal case, not only must the envelope of this high frequency carrier wave be an exact replica of the original signal, but it should reproduce changes of signal amplitude in a linear manner up to the maximum depth of modulation required. In the case of telephony, it has been shown that this involves the transmission of a wide band of audio frequencies; and hence for a given intensity of signal the aerial current change should be the same throughout this band, if distortion is to be avoided.

The degree to which these conditions are fulfilled in any transmitter depends very much upon the services required of it. For instance, broadcasting, since its primary object is entertainment, requires a high degree of linearity over a very wide frequency band, 30 to 8,000 cycles, up to large depths of modulation, 90%; but, on the other hand, the power efficiency is of secondary importance and the simplicity of circuit and prime cost of little interest. The case of transmitters handling commercial telephony is different; here the frequency band can be limited from 250 to 2,750 cycles, the frequency response curve need not be strictly linear, but the power efficiency is of the greatest importance. Although modulation circuits both for broadcasting and general purpose sets are the same in principle, the greater accuracy of response required in the first case necessitates much more elaborate apparatus.

Amplitude control of a carrier can be made, either directly upon the high frequency wave, or on the direct current supply before conversion into high frequency. Certain systems of control are suitable both for self-oscillators or driven sets, and others suitable only for driven transmitters. In this connection it is well to point out that on short waves it is quite impossible to transmit intelligible speech with the self-oscillator type of transmitter (however efficiently it is modulated), except over the very shortest distances and with wavelengths above 50 metres. The reason for this is due to multiple echo effects giving more serious distortion when the transmitted wave is not kept extremely constant, and in consequence we shall confine our discussion to systems suitable for the driven type of transmitter. It is important to observe that all methods involve the setting of the system to some asymmetric condition; for modulation, like detection, cannot be carried out unless the circuit is so adjusted.

The general methods of amplitude modulation suitable for a driven short-wave transmitter may be grouped under the following headings:

(1) Anode Modulation:
 (*a*) Choke.
 (*b*) Series.

(2) Low Power Modulation.

(3) Grid Bias Modulation:
 (*a*) Current.
 (*b*) Voltage.

(4) Cathode Modulation.

(5) Suppressor Grid Modulation.

We do not propose to deal with the high-efficiency, phase modulation systems as developed by Chireix, Doherty, and others, as these are only successful at long and medium wavelength.

Anode modulation is one of the simplest methods employed and involves the change of amplitude of supply voltage to the anode of the H.F. transmitting stage.

Consider say a stage of a high-frequency transmitter being operated under Class B conditions, having an anode supply

voltage of E. The load line will roughly be as indicated by ABC Fig. 248, and we can assume the conversion efficiency would be approximately 65%, this giving an output high frequency current I, Fig. 249. If now we raise or lower the D.C. voltage, still keeping the valve under Class B conditions, the load line will slide parallel to ABC, the conversion efficiency will not materially change and in consequence we shall obtain a linear relationship of output H.F. current to anode voltage as shown in Fig. 249, provided we have sufficient peak emission in the valve.

If a sinusoidal change of anode voltage is provided from a modulator system, the relationship of anode voltage and

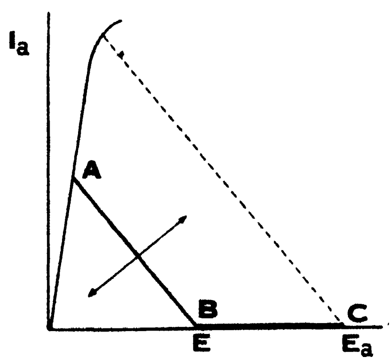


FIGURE 248.

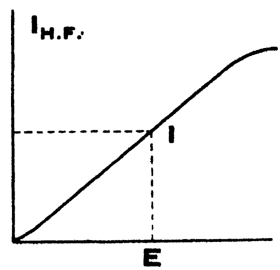


FIGURE 249.

current will be as shown pictorially in Fig. 250. Observe that if the conversion efficiency remains constant, the power supplied and converted surges from high to low levels each modulation half-cycle. At the peak of modulation, the instantaneous voltage and current are doubled, hence the power is quadrupled. If the carrier is modulated 100%, then the average power rises to a value of 1·5 times the power in the carrier condition, and there will be in consequence a rise of H.F. current during such modulation of $\sqrt{1{\cdot}5}$. This is obvious since the power due to carrier and side-bands is 1·5 the carrier power and this must be equal to $I^2_{R.M.S.} \times R$. The rise of H.F. current for this and different percentage modulations is shown in Fig. 251.

Since the dissipation in the amplifier valve under modulating conditions must rise by 1·5 times for sine modulation, and more for broader waveforms, this fact, and the greater peak emission

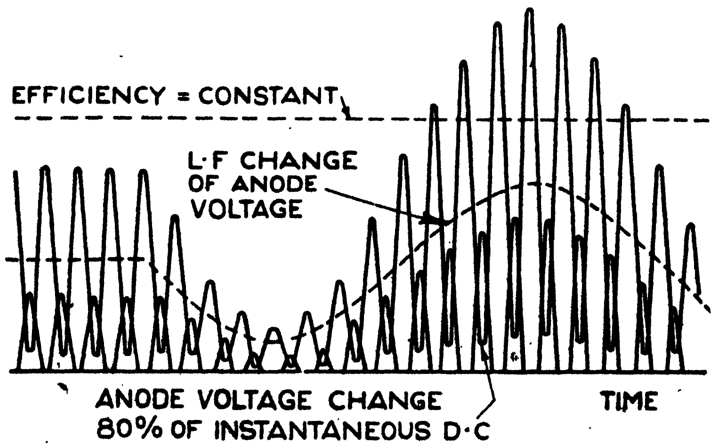


FIGURE 250.

required, must be kept in view when choosing a suitable valve. Or conversely, for a valve designed economically for telegraph conditions, we must, if we desire to modulate it, reduce the

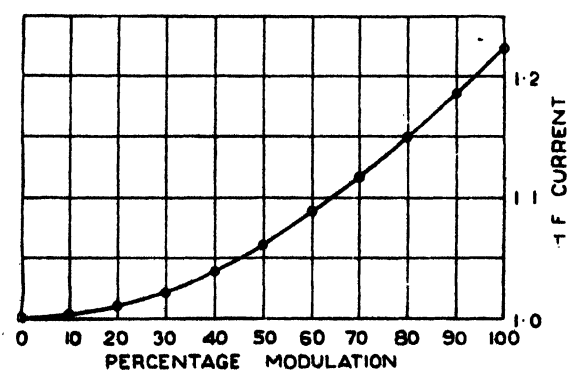


FIGURE 251.

carrier level by an amount which depends upon the percentage modulation required, thus:

$$\text{Modulated Power Possible} \Big\} = \frac{\text{Unmodulated}}{\text{Power}} \times \left(\frac{1}{1+\frac{K^2}{2}}\right)$$

Anode Modulation by Choke Feed. One method of obtaining a variation of anode voltage is by the use of an iron core choke in the supply. This method, which is due to Heising, is sometimes referred to as a constant-current control, and is suitable either for a self-oscillator or driven valve.

Since modulating signals are usually all within the audio frequency band, the most direct method of control is to include an iron core transformer in the main supply, any signal across the primary producing voltages across the secondary which must be commensurate with the D.C. voltage to be controlled.

The design of a satisfactory transformer for such purposes has really only become practical with the advent of the push-

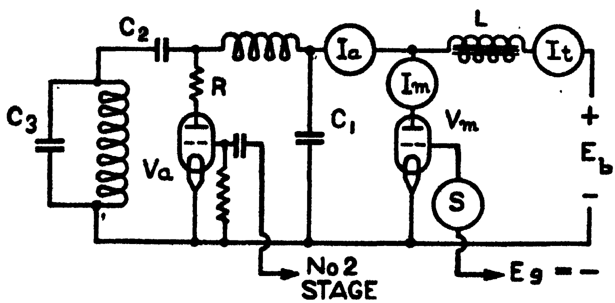


FIGURE 252.

pull modulator arrangement, as described on page 403, and since a transformer is not essential, modulation is often effected by the use of an iron core choke in the main supply lead, with the modulator valve in shunt as shown in Fig. 252.

Let us consider the choke fed arrangement where V_m is the modulating valve, V_a the H.F. amplifier being driven at a 70% efficiency, say, and L an iron core choke, whose reactance can, for the moment, be assumed to be infinity to the signal frequency. We have to show, first, how modulation impressed on the grid of V_m controls the amplifier in the manner previously discussed, and, secondly, what adjustments are necessary to obtain the greatest efficiency, or linearity.

Since the modulating valve V_m may have to control a power as great as the carrier under certain conditions, it will be fair to make the assumption that the modulator feed I_m should

at least equal the feed to the amplifier I_a, in the quiescent condition.

The total power supplied is thus double, and is divided equally between amplifier and modulator valves, but whereas the former converts 70% into high frequency carrier, the latter dissipates the total power, and must therefore have a dissipation rating three times that of the amplifier. This static condition of the modulator is shown at "*O*," Fig. 253, which represents

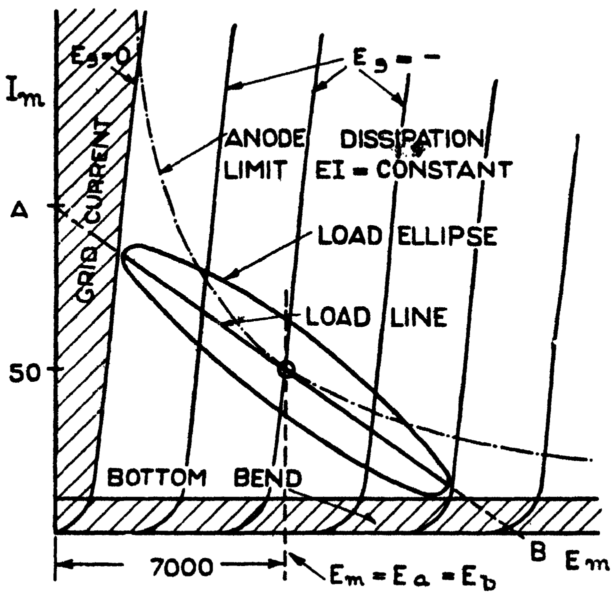


FIGURE 253.

an E_aI_a curve of a valve which will fulfil the required conditions.

If now a low frequency sine signal is impressed on the grid of V_m, the current flow resulting will depend upon the circuits external to the modulator, and we can obtain an idea of the results by combining the modulator valve and load characteristics.

The circuit is to be considered as shown in Fig. 254, where the valve represents a source of low frequency, across which is an iron core choke L, and by-pass condensers C_1, C_2 and C_3, all of high reactance; and the modulated amplifier circuit,

which may be considered as a non-inductive load to low frequency of value

$$R_a = \frac{E_a}{I_a}$$

Assuming for the moment the shunt reactance of choke and condensers to be very large compared with R_a, they can be neglected, and the load characteristic is, therefore, a straight line of slope $\frac{I_a}{E_a}$. For the numerical example previously considered, this is $\frac{50}{7{,}000}$, or as AB, Fig. 253, the length of this line being dependent upon the excursion of modulator grid, and its position such that it passes through "O," the average condition.

The action of the modulator now becomes clear, for change of modulation grid potential by the sine signal leads to in-phase current changes, and antiphase anode voltage changes, which if the characteristic was ideal would swing the anode potential of the modulator, and hence of the amplifier, between the limits of twice the static E_m and zero, and swing the modulator current between zero and twice I_m.

This changing modulator current cannot be satisfied by the supply, because of the choke L, which prevents current change, and hence the variations of I_m are interchanged with those of I_a, the latter rising when I_m falls, and vice versa.

Thus, by impressing a low frequency signal on the modulator grid, the supply voltage to the amplifier is varied in accordance with the signal, and consequent modulation of carrier results, the additional power necessary to effect this being forthcoming from the modulator itself. Because of the antiphase condition of E_m and I_m the power dissipated during sine modulation drops to half that of the quiescent condition.

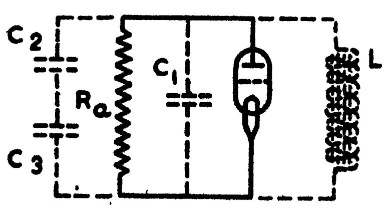


FIGURE 254.

Another way of considering the matter is that because the choke prevents changes of feed from the supply, the total current I_t is constant. Hence any change of feed through I_m

is accompanied by an opposite change of voltage E_m and amplifier feed I_a; and because E_m and I_m are in opposite phase, and E_m and I_a in phase, the amplifier gains power and the modulator loses.

In the following table the power relation of modulator and amplifier are shown both for quiescent and modulating conditions. It is necessary to understand the significance of these figures, as although the distribution of power changes when modulating, no indication of this is given by the feed instruments I_t, I_m and I_a, which remain stationary throughout.

	Carrier.		*Sine Modulation* 100%.	
	Average.	Peak.	Average.	Peak.
E_b	*7,000		*7,000	
I_t	*100		*100	
I_a	*50	300	*50	600
I_m	*50	—	*50	100
E_a	7,000	12,600 $(E_b + \cdot 8E_b)$	7,000	25,200 $(E_b + \cdot 8E_b)$
E_m	7,000		7,000	14,000
W_t	700		700	
W_a	350 < 105		525 < 157·5	
W_{HF}	350 < 245		525 < 367·5	
Conversion Efficiency	70%		70%	
W_m	350		175	
I_{HF}	*1·0 R.M.S.		*1·23 R.M.S.	

* Instrument measurements.

NOTE.—The peak current values are based on the assumption of Class "C" working with $\dfrac{I_p}{I_a} = 6/1$.

Observe that the only instrument that gives any clue to the effective modulation is the H.F. ammeter, and this rises only from 1·0 to 1·23 for 100% modulation. For percentage modulation of less amount the rise is as shown in Fig. 251.

Of course, peak voltmeters across either the speech choke L or the valves will give indication of control, but since it is H.F. output we are interested in, the change of I_{HF} is the most satisfactory guide to the performance of the system.

Obviously a modulator cannot have ideal characteristics to allow 100% modulation, limitations being caused both by bottom bend of the characteristic and by grid current. It is evident that in order to get even a medium modulation percentage we must use a very low resistance modulator valve, such that the $E_a I_a$ curve for $E_g = 0$ is very steep; but with the best possible design, 80% represents a maximum value for modulation with a straight circuit, although a modified arrangement can be made to give 100%, as will be described later.

The excursion of grid voltage should be such that the peak of modulation does not run into either region mentioned, a sufficiently low resistance valve being chosen to require a large negative bias in the quiescent condition.

Where a large band of frequencies is to be signalled the straight line load characteristic does not represent the true load on the modulator at all frequencies; for at low frequencies (50 cycles) the reactance of L is commensurate with, or may even be small compared with, R_a, whereas at higher frequencies (5,000 upwards) the reactance of the shunting condensers will be small. In consequence, at the top and lower end of the frequency band the load curve is no longer a straight line, but opens out into an ellipse, as shown in Fig. 253. This means the percentage modulation will fall away at these frequencies, and unless the average feed is raised to accommodate the ellipse, distortion will result.

Chokes are suitable only for sets of low power, as the D.C. component of feed necessitates a very large iron section if saturation of core is to be avoided. This may be overcome by the use of a tapped choke, usually 1/1 ratio, where the connections are such that the modulator and amplifier feeds pass in opposite directions; thus the magnetising effects of these feeds cancel if they are equal, but any variation of feed by the modulator impresses its voltage changes on the amplifier in a similar manner to that described in the choke circuit.

Earlier in the chapter it was mentioned that a modified circuit could be set up to give 100% control. If a resistance, shunted by a condenser to pass low frequency, is inserted in series with the amplifier D.C. circuit (R Fig. 252, the condenser not being shown) this will drop voltage of amount $I_a R$, where

I_a is the average feed. Thus the D.C. volts to be converted will not be E_m, but some lower value. For instance, in the example given where the feed is 50, with $E_m = 7{,}000$, if 40,000 ohms is included in series with E_a the voltage across the valve is now but 5,000, and thus the carrier current will fall to ·8 amp.

But the modulator valve still controls 7,000 volts; its load characteristic is not materially altered by the 40,000 ohms, and in consequence a 75% control by the modulator (5,000 volts) creates 100% control of the reduced amplifier volts. This method of achieving full control is, of course, at the expense of efficiency and output.

If the transmitter is large, to obtain sufficient swing on the grids of the modulator valves may necessitate amplification on the original audio frequency signal, be it from a microphone or an incoming line, and it is usual to adopt a multi-stage resistance capacity type of amplifier with a correcting network, if linear response is desired over a wide band of frequencies.

Class B Push-Pull. By utilising two valves in push-pull and operating them at a point just above cut-off, a system is provided in which the D.C. loss in the modulators can be avoided. Thus instead of the modulator being connected if shunt with a choke, the two modulators are connected in push-pull across the primary of a transformer, the secondary of which is in series with the H.T. supply to the main unit to be modulated, i.e. the modulated amplifier, the circuit diagram of a modulation unit being as shown in Fig. 255. To avoid passing the D.C. feed to the modulated amplifier through the transformer secondary it is usual to shunt-feed the latter through a choke as indicated. Such a circuit is often supplied with negative feedback to correct for distortion which may be introduced by the main transformer, Fig. 255 showing a feedback circuit of simple character.* In this circuit the feedback voltage is derived from a resistance unit R_1, in series with the main transformer secondary, the voltage from which is injected back into the primary of the line-to-grid transformer of the first audio stage. To ensure stability of working and to provide a certain amount of audio frequency discrimination a resistance-capacity choke network, RCL, is connected across the feedback resistance.

A trouble associated with early push-pull modulators was that due to excessive voltages being developed across the transformer during over-modulation periods. This can be prevented by the fitting of a limiting device, usually a neon tube, across a convenient modulator valve.

Series Modulation. The series modulation circuit has been developed to operate in one of two ways. In the first type the modulation is applied between grid and cathode of the modulator valve, and in the second the modulation is applied in such a way that it includes the load circuit as well

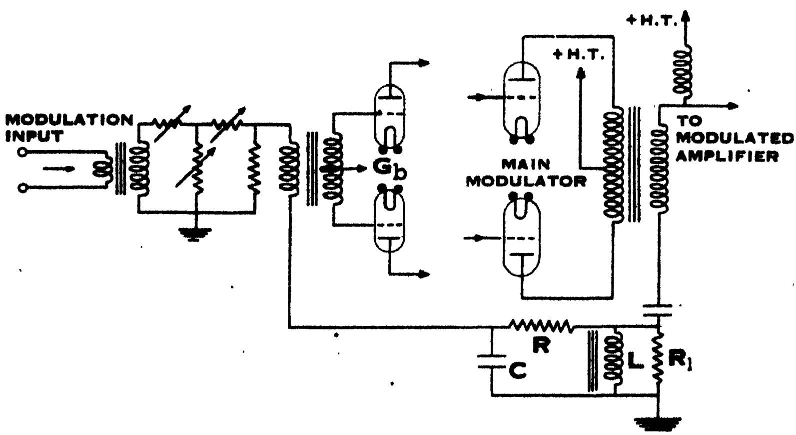


FIGURE 255.

as the grid-cathode circuit of the modulator valve, in which case the circuit is of the type known as a " cathode-follower."

Dealing with the first case, where the modulation is applied between grid and cathode, it is immaterial whether we arrange for one side of the H.F. circuit, or one side of the modulator, to be at earth potential. If the former, we shall have the problem of feeding in the modulation at a H.F. potential above earth, whilst in the later we have to arrange an H.F. circuit which is all at the modulating voltage above earth. Both circuits have been used, although the latter is probably a rather more simple problem, and Fig. 256 shows a series modulation circuit with modulator at earth potential.[3]

The circuit may be considered as consisting of two resistances in series, the amplifier resistance R_a being of fixed

value, and equal to $\frac{E_a}{I_a}$; and the modulator resistance R_m, which is variable. It is clear that if R_m can be varied between values of zero and infinity, the anode voltage to R_a will vary between E_b, the supply voltage, and zero, but as with the choke method of modulation, limitations of the characteristics prevent this range being achieved.

In analysing this circuit we can follow exactly the same procedure as in the choke modulation case, by setting up the

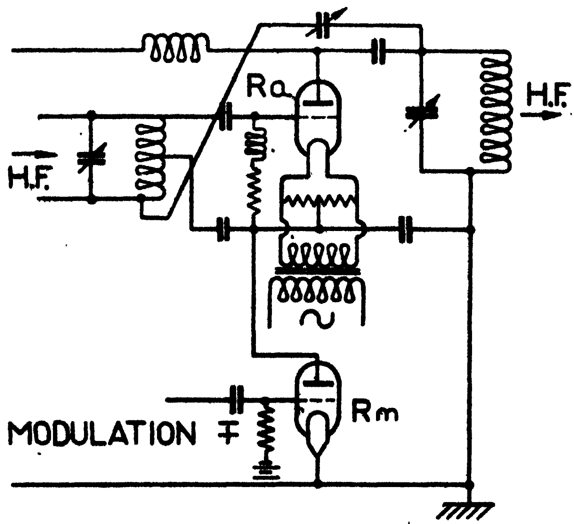


FIGURE 256.

E_aI_a characteristics of the modulator valve, and laying across these the load line due to the H.F. amplifier.

As an example let us use the same modulator valve and high frequency circuit as in the choke modulation circuit just discussed. The high frequency circuit forms the load in series with the modulator and its value $R_a = \frac{E_a}{I_a} = \frac{7000}{50}$ giving a load line of slope AB. Then if we had an initial supply E_b of 14,000 volts (instead of 7,000) the load curve would be BOA (Fig. 253) and a bias on the modulator to the point O would determine the quiescent condition.

As before, we have 7,000 volts across both modulator and amplifier, and the same current through each, but since they are

in series, the supply voltage is not E_b, as marked in Fig. 253, but 14,000, the point B. The power supplied is the same as before, the supply voltage being doubled and the current halved. Half this power is dissipated by the modulator and half delivered to the high frequency valve, the latter converting some 70% of its input into high frequency output and dissipating some 30%. Thus the same valve anode dissipation ratings as for the choke case will be necessary.

If now we consider the effect of an alteration of modulator grid voltage, this will sweep the modulator current in phase and the modulator voltage in antiphase along the load line and create amplifier anode voltage and current changes, the modulator losing and the amplifier gaining power by an amount dependent upon the depth of modulation and type of signal. Thus the action is similar to the choke modulated case.

It is clear we shall have the same limits due to bottom bend and grid current and in consequence the maximum change of modulator voltage will be less than 7,000, and it is not possible therefore to modulate to 100%. In the case shown in Fig. 253, that is with equal voltage across modulator and high frequency valve in the quiescent condition, the modulation factor is about 70%. If, however, the initial quiescent condition is arranged such that the modulator takes greater voltage than the high frequency valve, the modulation percentage can be raised to well over 90%.

There is no iron core choke in the series circuit and the load line, therefore, is always substantially a straight line, and it is found possible to obtain modulation over a greater length of characteristic without introducing distortion. Measurements made with this type of control show that the distortion factor [1] can be kept within 3% even up to a modulation percentage of 90%, and in consequence the series modulation method is very suitable where high quality is required, or where the modulation involves a wide frequency spectrum.

[1] The distortion factor is defined as follows: If a pure sine modulating tone is applied to the apparatus and the output modulated wave analysed, then the percentage distortion factor is given by

$$\frac{\text{Root sum square of harmonic amplitudes}}{\text{fundamental amplitude}} \times 100$$

Series Modulation, Cathode-Follower System.[4, 5, 6] We will now deal with the cathode-follower series modulation circuit, in which the modulator valve is at a H.F. potential to earth, one side of the H.F. circuit is earthed, and the modulating voltage applied between the grid of the modulator and earth and not between grid and cathode ; thus the modulating voltage is also across the load circuit. A simplified diagram of such a system is shown in Fig. 257, where V_1 is the series

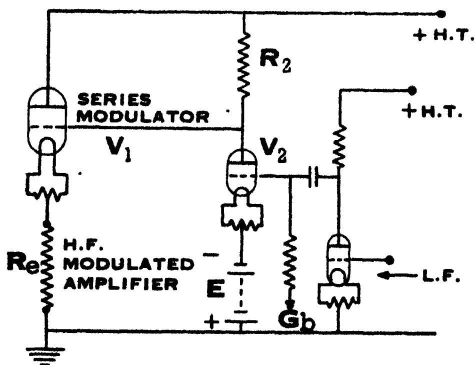


FIGURE 257.

modulator, and R_e represents the H.F. amplifier circuit whose equivalent resistance will be equal to a value

$$R_e = \frac{E_{av}}{I_{av}}.$$

The grid of the series modulator is fed from the anode of the valve V_2, in the anode circuit of which is a resistance R_2, the value of this being made appropriate to supply the correct voltage to the grid of V_1. If the series modulator stage consists of more than one valve in parallel, then the grid of each valve will be tapped separately along this resistance so that each valve shares the load equally. Between the cathode of V_2 and earth is included a D.C. supply with positive to earth so as to increase the H.T. potential on this valve, and so allow V_1 to modulate the main anode potential 100% without running into grid current limitations. The grid of V_2 is fed from an orthodox resistance-capacity amplifier between grid and earth.

Examination of Fig. 257 indicates that the currents in the valves V_1 and V_2 will be in opposite phase, for when the current in V_2 is cut off, the voltage applied to the grid of V_1 will be the least negative (relative to its cathode), whereas a rising current through V_2 will cause a fall of potential on the anode

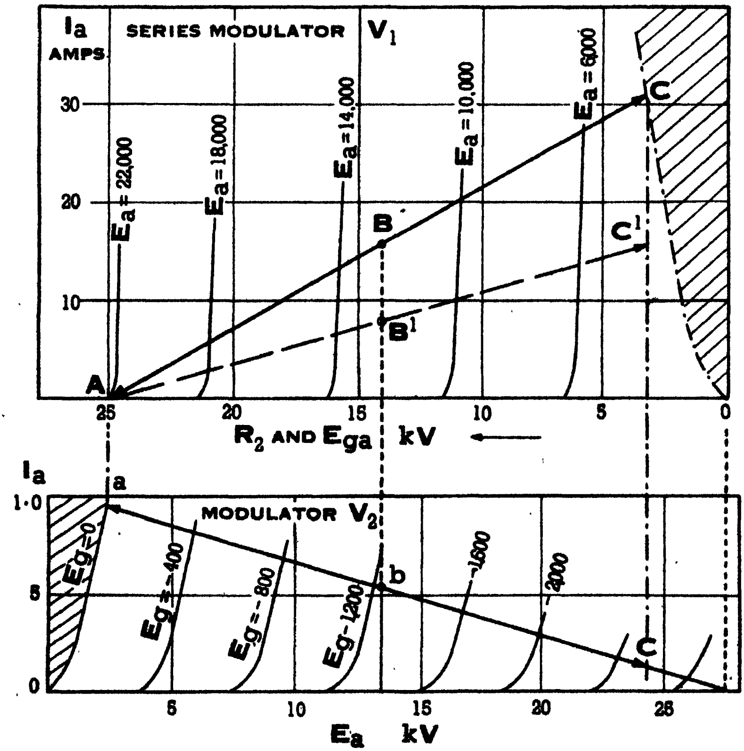


FIGURE 258.

of V_2 and in consequence an increasing negative on the grid of V_1.

In order to explain the working of the circuit clearly a definite case has been set out, the figures being taken from the Marconi 150 kW short wave broadcast transmitter as supplied to the B.B.C.

In this case the carrier power to be modulated is 150 kW, and if the amplifier H.T. is to be 9,500 volts, the carrier current will be 15·7 amps. In order, therefore, to obtain

100% modulation without distortion we have to arrange the series modulator so that we can obtain a peak current of 31·4 amps without running into grid current. When the current is cut off, the modulator V_2 (which will then be passing full current) must not take grid current either. How this is accomplished can be seen by examining Fig. 258 where the characteristics of both valves V_1 and V_2 have been aligned one above the other so that the load lines can be directly correlated. From this figure it is observed that whereas the characteristics of V_2 are plotted in the usual way, namely as E_aI_a curves, those of V_1 have been plotted to show the relationship between the anode current and the voltage between anode and grid, instead of anode-cathode. The reason for plotting them in this way, is that the voltage change between grid and anode is the same as that across the resistance R_2 and hence can be directly related to the curves of V_2; whereas because of the resistance load included in the cathode circuit and the fact that the grid potential is across grid and earth, the relation between I_a and E_a in V_1 is not so directly obvious.

Examination of these curves shows that the positioning of the curves of V_1 is such that the origin of voltage of V_1 is exactly above the current cut-off point of V_2, since at this point there is no voltage drop across R_2, and in consequence across E_{ga}. We cannot work down to this point, however, but only to such a point that no grid current flows in V_1 (the area shown shaded), indicated at *Cc*, this point being the peak of modulation. The limit at the other end indicated by *Aa* will be set by the $E_g = 0$ curve of the valve V_2, the excess voltage provided by E being sufficient so that the anode current of V_1 can vary from 31·4A to zero, without running into grid current on either valve. The carrier setting will be indicated by a point half-way between these limits, shown at *Bb*.

In order to indicate the actual voltages across the various parts of the circuit for the peak, carrier, and trough conditions, figures have been extracted from these characteristics and set out in diagrammatic form in Figs. 259 and 260. In Fig. 259 the figures placed between arrows indicate difference of potentials, and figures by themselves indicate potentials above earth. It can be seen that the potentials on the modulator V_2 exceed those on the series modulator V_1, although of course there is a

power gain from the V_2 stage to that of V_1. Figure 260 sets out the variation of voltage and current with time for the same peak conditions, it being assumed that a sinusoidal voltage is

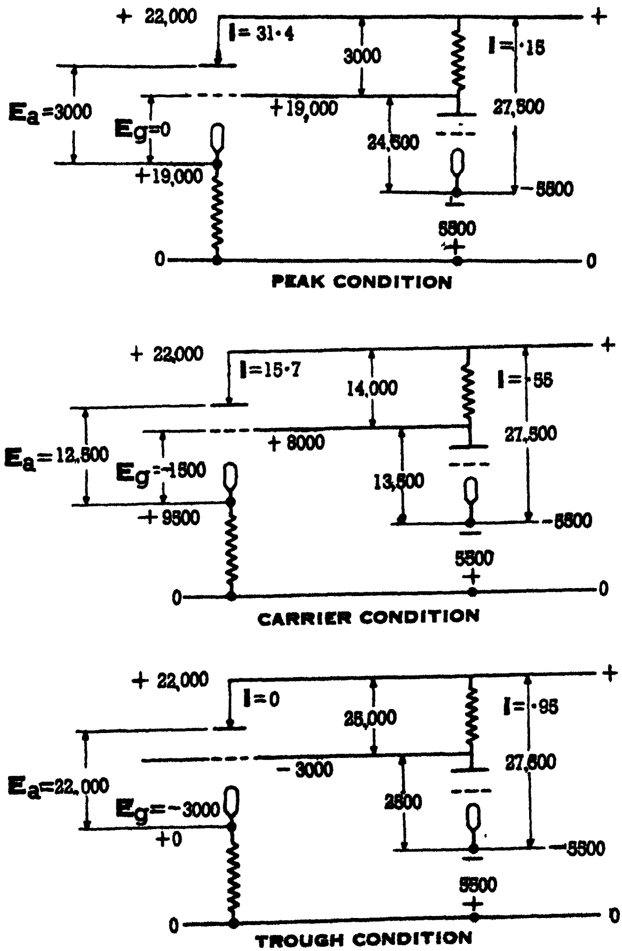


FIGURE 259.

applied to the grid of the modulator. These curves show clearly that the anode of the series modulator remains at a constant potential above earth and that the other potentials vary relatively to it, and to earth. The voltage changes, grid-cathode, grid-earth, cathode-earth all vary in phase with one

another and with the change of I_a. Of course, from the conversion point of view this phase relationship is the same as that of an ordinary circuit, as the anode voltage and anode current are virtually in anti-phase, if we consider the cathode as a datum. It will be noticed that the changes of cathode potential are the same as those of the grid potential (relative to earth) both in phase and amplitude (nearly), hence the name cathode-follower. They would be the same in amplitude if a very high-μ valve was used.

The inclusion of the load resistance in the grid circuit means that there is no voltage magnification, but on the contrary a

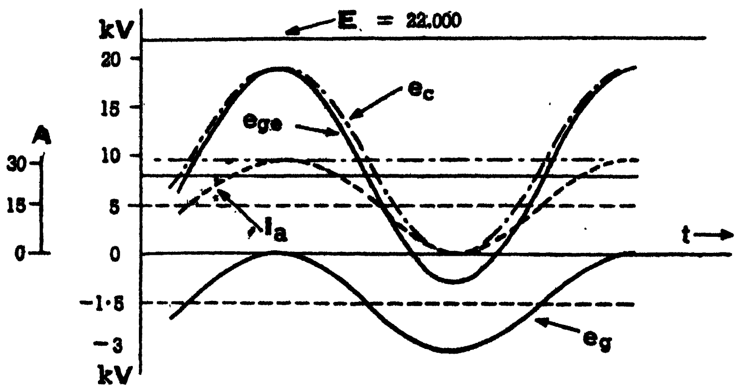


FIGURE 260.

small reduction; it is, therefore, equivalent to a 100% negative-feedback circuit, and as such it has a number of advantages where high linearity is desired, amongst the advantages being the following:

(1) With a linear grid swing, non-linearity of the series modulating valves will have a negligible effect.

(2) With a non-linear load the voltage swings are still linear.

(3) A variation of load resistance over wide limits will not affect the modulator setting appreciably. This can be seen by considering Fig. 258 where the load line for a load of double the resistance value, shown at $AB'C'$ has been added.

Other advantages are that A.C. lighting for the series modulator can be used without any hum ripple appearing,

variations due to H.T. changes are negligible, and with paralleled modulators the removal of one valve will not alter the operation of the circuit. Actually it is probable that the future of such circuits will be more for low-power, high-linearity circuits than for high power.

The foregoing discussions have been confined to the modulation of any single stage by anode modulation of that stage, it being assumed that the grid of the H.F. stage was supplied from a constant amplitude H.F. source. If such modulation is operative on the final power stage, the system is known as a high-power, modulation system. Observe that such a high power system will involve the design of an unmodulated H.F. amplifier and a low-frequency amplifier capable of handling, in its final stage, a power equal to that of the H.F. amplifier it is to modulate.

We may, however, prefer to modulate at some quite low-level point in the system. In this case the problem is different, as in all stages subsequent to that of the modulated amplifier we require to amplify a H.F. modulated wave. In essence the problem is similar to that of a receiving amplifier except that we must keep in mind the fact that we have power to deliver, and not voltage as with the receiver, and therefore conversion efficiencies of power stages will be important. Such a system is called a low-power modulation system.

Low Power Modulation. Consider a transmitter such as is shown in Fig. 168 (Chapter X). If the input-output characteristic of any amplifier stage is considered, say stage No. 2, it is observed that input H.F. driving volts and output H.F. current have a relation, as shown in Fig. 171 (Chapter X), indicating a linear connection up to the full driving conditions. The curves for I_{HF} and efficiency against grid voltage swing are shown again in Fig. 261 (and 264). Thus, if we set such an amplifier to the point A statically (Fig. 261), thus making it a Class B amplifier, and drive it up to the point "O" by high frequency impressed on the grid (representing the carrier condition), any modulation impressed on the input high frequency will lead to modulated high frequency in the output, and if the input frequency is modulated between limits of OA, OB, we shall get a high frequency output 100% modulated.

In the carrier condition, namely O, the efficiency of the amplifier stage unmodulated is but half the efficiency under full driving condition. Hence modulation is achieved by

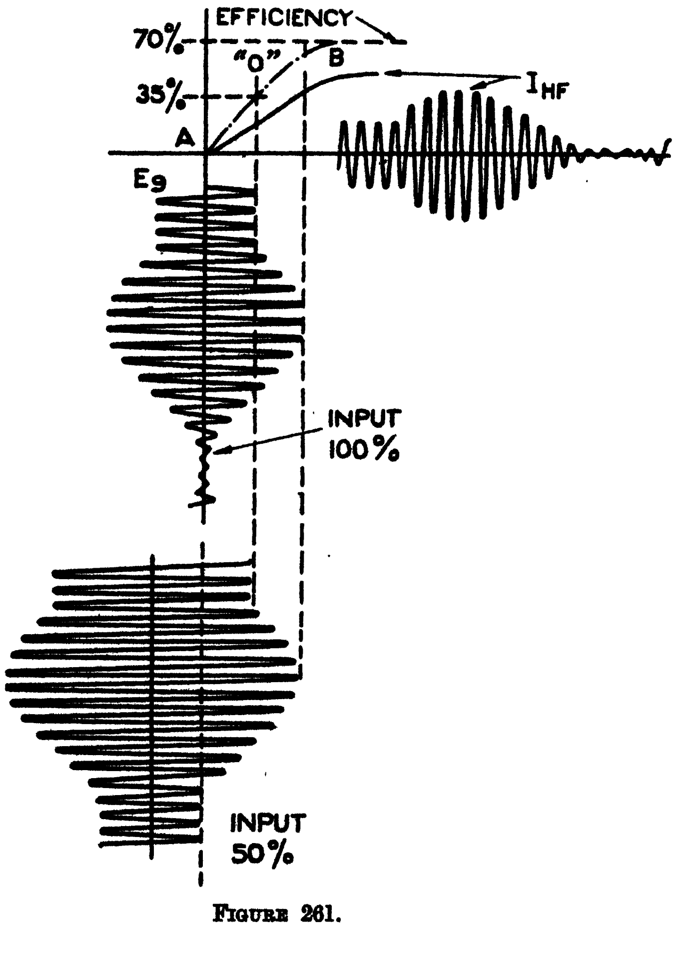


FIGURE 261.

variation, not of anode voltage, but amplifier efficiency, a point which cannot be stressed too much.

This is shown in Fig. 262 and 263, where the E_aI_a curves of the amplifier are shown, and the cyclic change of voltage, current and efficiency, for sine modulation, Fig. 261 showing the connection between the input modulated wave and output. Observe that at the peak conditions of modulation the anode

voltage and peak emission are those normal to a telegraph set driven at 70% efficiency.

A second point to observe is that the depth of modulation of one stage has no relation to that of the other. The input may be deeply modulated, and yet no modulation appear in the output. This would occur, for instance, if the stage was set too high in the carrier condition, as *B*, above.

On the other hand, it is possible to obtain a 100% modulation of output even if the input has but shallow modulation. For instance if the amplifier is biassed very negative, say beyond the point *A* as shown, thus making it a Class C amplifier, and the input H.F. is increased to such an amplitude that it still drives the amplifier up to the same point "*O*" in the

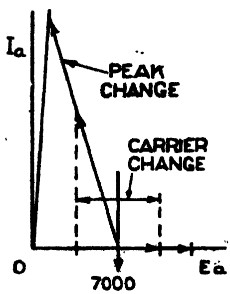


FIGURE 262.

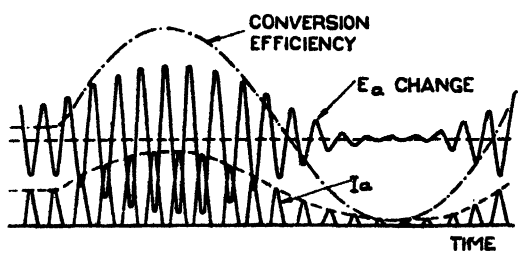


FIGURE 263.

carrier condition, then only moderate control of this larger input (50% in the case shown) will still give full modulation of the amplifier output as shown in Fig. 261.

To obtain the input H.F. wave modulated by the signal, we still need, of course, some form of control system on the previous stage, but, as explained, we need not employ 100% control at this stage.

There appears to exist erroneous ideas concerning the overall efficiency of low power systems. A comparison deduced from first principles indicates that both systems have equal power efficiency.

Referring to Figs. 261 and 264 it is clear that we can control the output in one of two ways:

By varying the grid excitation voltage in accordance with the modulation, keeping the grid bias constant; or by keeping

the grid excitation voltage constant and varying the grid bias at the modulation rate as shown in Fig. 264. The former method has already been described, and the second method which we will now discuss can be carried out in one of two ways.

Low Power Grid-Current Modulation. In this arrangement the grid bias changes are produced by introducing as a grid resistance a valve as shown in Fig. 265. By varying the resistance of this "valve" leak at the modulation

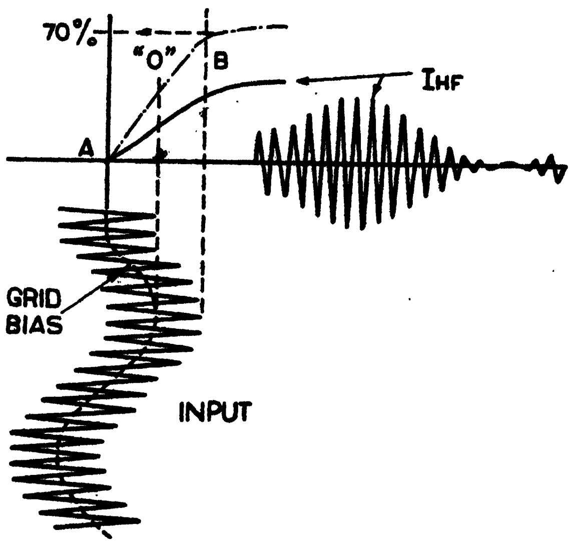


FIGURE 264.

rate, the bias on the stage to be modulated will be changed. It is not a simple matter to obtain correct linearity by this method, but it is useful for certain classes of work and it has not the frequency limitation of the choke methods.

Low Power Grid-Voltage Modulation. D.C. grid voltage control is but slightly different. Instead of changing the grid bias by varying the resistance value in the grid circuit, the modulation is impressed across a grid circuit resistance, the voltage drop across the resistance due to modulation controlling the grid bias at the modulation frequency.

In Fig. 266, *A* and *B* represent the valves of the last stage of a high frequency amplifier. Included in the D.C. grid

circuits of these valves is a resistance, R, which is of low value, such that the normal grid current of the two valves produces a negligible negative static bias. Across the resistance, R, is placed the modulator valve, M, and its anode supply voltage, both carefully insulated from earth. It will be clear that if the anode current of the modulator is large compared with the grid current of A and B, the voltage drop produced will be a function of the modulator valve feed and in consequence the negative

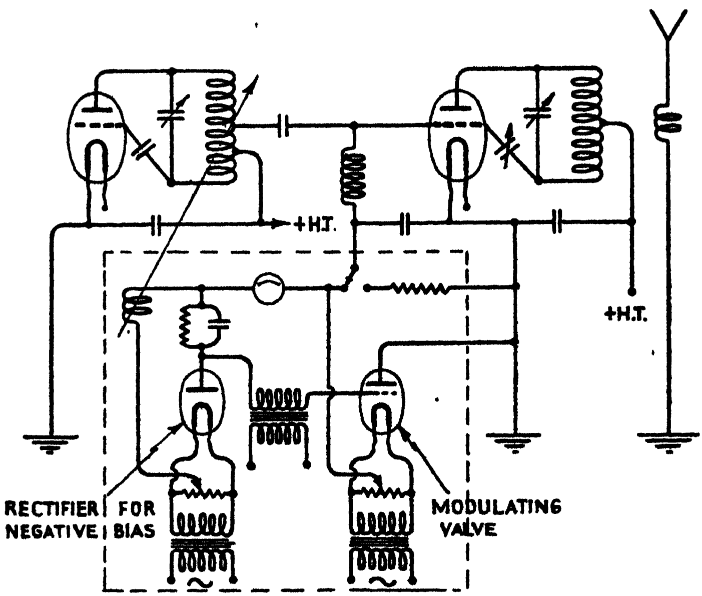


Figure 265.

bias produced on the H.F. valves will follow directly, changes of modulator feed.

Thus if the modulator is set to the centre of its characteristic in the quiescent condition, application of the signal voltage to the modulator will create similar changes of amplifier grid negative bias. Provided, therefore, the initial bias is such as to set the H.F. amplifier to the centre point of its output curve as previously explained, linear change of H.F. output will result.

This system has been very successful in dealing with high modulation frequencies and has been used for a television transmitter having a frequency response up to two megacycles

Cathode Modulation.[6] A popular system of modulation for small power sets, is the so-called "Cathode-Modulation" system, where the modulating E.M.F. is injected between the cathode and earth, the grid and anode circuits being returned to the earth side of the modulation system. The circuit of a cathode-modulated stage is shown in Fig. 267, where included in a conventional driven stage is a modulation transformer "M" with its secondary circuit connected between the cathode of the amplifier valve and earth, the primary being supplied from a normal push-pull modulator.

Assuming the H.F. stage is being driven from a constant-amplitude source in the usual way, then the application of

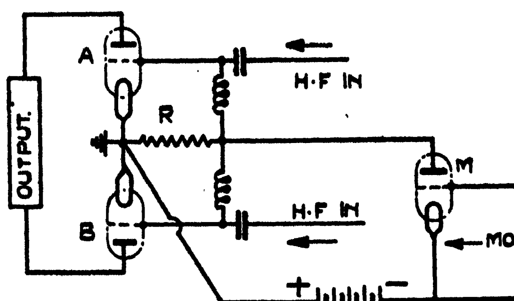


FIGURE 266.

an A.C. modulation potential across the secondary of *M* will produce a voltage which will vary the cathode potential about earth. Or, if we consider the cathode as a reference point, then the common point of connection of grid and anode (namely *E*) will vary at low frequency as regards the cathode, so that as the anode voltage increases above the D.C. value, relative to the cathode the grid will become less negative, and as the anode potential falls the grid potential will go more negative. Thus such a type of control is a combination of both grid and anode modulation, the amount of control being determined by the voltage supplied from the modulator and the value of grid bias resistance, R_g, Fig. 267, provided. That is to say, we can increase the proportion of anode modulation relative to grid modulation by increasing the modulation voltage (and the size of the modulator); and

making R_g, and in consequence negative bias, large. By reducing the grid bias a smaller applied modulating-voltage will have a greater control of the grid potential and the system becomes nearer to a low-level type.

Claims have been made that it is possible to obtain a conversion efficiency as high as 50% with 100% modulation, and although it appears fairly easy to set up a circuit to give such a

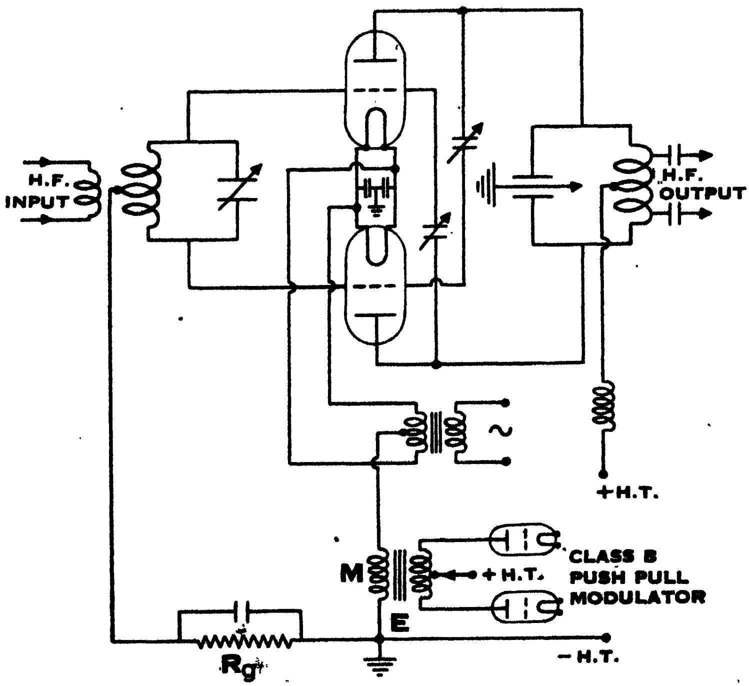


FIGURE 267.

figure, examination of the wave form will indicate a considerable harmonic content. Some experimental figures obtained at the School of Wireless Communication at Chelmsford, indicated that to obtain a circuit with less than 2% harmonic content, it was not possible to get a conversion efficiency greater than 35% for 100% modulation although of course, if the modulation percentage is reduced, and the circuit readjusted, the conversion efficiency can be increased. Fig. 268 shows that a very linear relationship between modulator input

and percentage modulation can be obtained, and it is found that under the maximum percentage-modulation condition, the modulation input power is but 10% of the output power. A small point to be observed is that the driving voltage needs to be about 50% of that necessary to drive the same amplifier stage as a Class C amplifier to full output, although the grid bias required is greater, being some six times the cut-off bias, as against the more normal value of 3·5. This means we are virtually operating the amplifier as a much under-driven, Class C amplifier. The frequency response of the

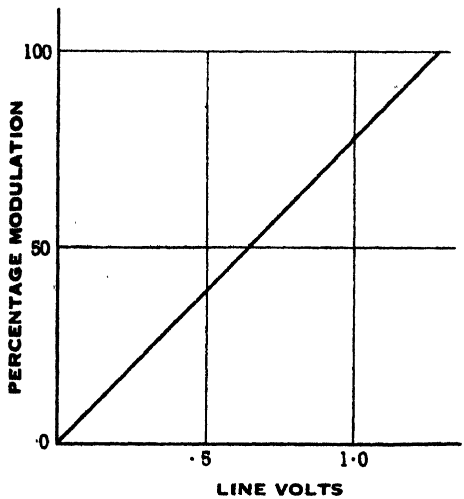


FIGURE 268.

circuit is good, being within 1 db. over a frequency range from 100 to 10,000 cycles, without any special precautions being taken, although it is found the load coupling is fairly critical.

Suppressor-Grid Modulation for Pentode Amplifier. On page 292 it was mentioned that one of the advantages of the pentode amplifying valve over the triode was the ability to use a simple system of low-power modulation. This is carried out quite simply by the application of the modulation to the suppressor grid$_{g3}$. Figure 269 shows the relationship between suppressor-grid volts E_{g3} and output current $I_{H.F.}$ from which it is seen that with the suppressor grid somewhat positive full output is obtained at a high conversion efficiency, and as

the suppressor grid is made negative both output current and efficiency fall, there being a linear relationship between E_{g3} and $I_{H.F.}$ over a considerable portion of the curve.

Thus if we adjust the value of E_{g3} to say -65 volts for the carrier condition, we can sweep to zero volts for the peak of modulation and to -130 volts for trough, and obtain a linear relationship throughout. This can be accomplished by including in the suppressor grid circuit, see Fig. 178, page 295, a modulation transformer and the necessary bias supply in series.

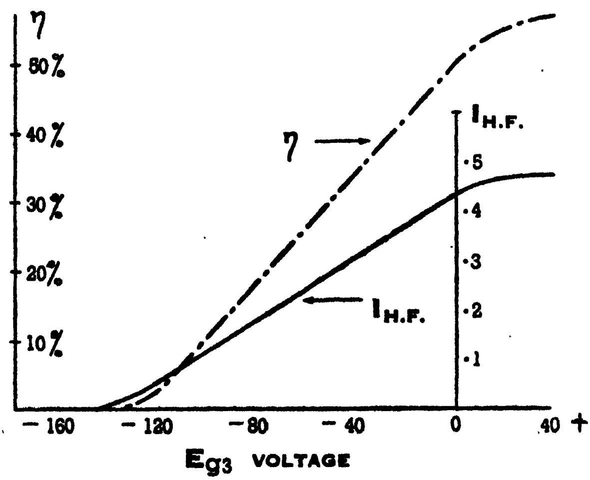


FIGURE 269.

The Negative Feed-Back Amplifier. The properties of an amplifier in which the output is coupled back to the input through an attenuating network, the feed-back voltage being in phase opposition to the input, were first investigated by Black. This arrangement, although reducing the gain of the amplifier, has many uses in low and high-frequency amplification, for the reduction of distortion and the improvement of the circuit linearity and as it has come into extensive use in the modulated amplifiers of transmitters, it is proposed to explain the principle very briefly. For a fuller explanation references given at the end of the chapter should be consulted.

Consider the circuit shown in Fig. 270, consisting of an amplifier having a gain μ, back-coupled through an attenuating

network—so that $E_b = \beta E_o$ is fed back to the input. Both μ and β will, in general, be complex quantities since there will be a phase change in the amplifier and in the attenuating network. We propose limiting the problem to cases where μ and β are simple numbers, indicating that no phase change is involved and the feed-back voltage is either in phase, or in anti-phase with the input voltage, the latter condition being the one in which we have most interest.

Considering Fig. 270.

We have : $E_o = \mu E_i$

$E_i = E_s + E_b$

also $E_b = \beta E_o$.

Hence, $E_o = \mu (E_s + \beta E_o)$

$E_o (1 - \mu\beta) = \mu E_s$.

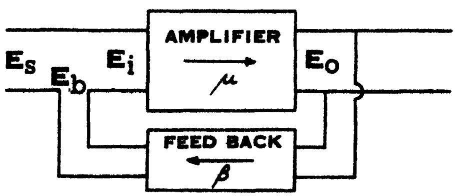


FIGURE 270.

The gain of the feed-back amplifier $= \dfrac{E_o}{E_s} = \dfrac{\mu}{1 - \mu\beta}$. (1)

If the phase of feed-back is positive, and the loss through the attenuator equals the gain of the amplifier $\left(\beta = \dfrac{1}{\mu}\right)$, the overall gain is infinity and self-oscillation will occur.

If the feed-back voltage is reversed β becomes negative and hence

the gain of a negative feed-back amplifier $= \dfrac{E_o}{E_s} = \dfrac{\mu}{1 + \mu\beta}$ (2)

Thus whatever the value of β, the overall gain must now always be less than μ, and if β is large, the overall gain is very small and largely independent of μ.

It is probably simpler to indicate the difference in action between a straight amplifier, A, and a negative feed-back amplifier, B, to do the same work, by taking an example.

Let us suppose we have a 10 V signal and wish to amplify this to 1000 V for the purpose of modulating a transmitter. The final stage of both A and B has to do the same work, and this stage introduces 10% distortion, giving an R.M.S. voltage of 100 V.

Considering A, the amplifier must be built to give a gain of 100, i.e. 20 dbs.

To provide the same output from the same input with the negative feed-back amplifier we shall need an increased gain which is dependent upon the value of β. Let us take a value of 0·009 for β. Then substitution in equation (2) shows that the amplification will need to be 1000.

The increased amplification necessary is clearly a disadvantage but this may be offset by some advantages now to be considered.

Suppose that changes in supplies cause the gain of both amplifiers to change by 10%. The output of A is clearly changed by 10%, namely to 1100 but that of B is given by

$$E_o = \frac{1100 \times 10}{1 + (0{\cdot}009 \times 1100)} = 1010\ V$$

so that only a 1% increase in output voltage has occurred. This stabilising property of negative feed-back is of great importance in measuring apparatus and in repeaters for telephone lines.

Let us now compare the distortion present in the outputs of A and B. For A this is 100 V. Let the value in B be D, then distortion voltage applied to input is $\cdot 009D$ and this gives $1000 \times \cdot 009D$ at the output stage, to be added to 100 V produced there, so that

$$D = 100 - 9\,D$$
$$\text{or } D = 10\ V.$$

Consequently, B produces much less distortion if the final stage is the same as in A. Alternatively, if a 10% distortion is permissible, it may be simpler and cheaper to build an amplifier which in itself produces greater distortion and apply feed-back to reduce it to 10%.

If "noise" is produced in the final stage (for example "hum" due to A.C. filament heating) then an argument similar to

that above will show that the output of noise from B will be 1/10th of that in A. In the case of noise or distortion produced in the input circuit of A or B, however, there will be no reduction in the output it produces in B. Noise or distortion introduced at intermediate stages in the amplifier, however, will give less output from B than A but the ratio will be less than one-tenth.

It will be appreciated that since μ and β are actually complex quantities, a given arrangement of feed-back which is negative over a range of frequencies may become positive for other frequencies. Careful design of the feed-back circuit is seen to be necessary, particularly when the amplifier has several stages.

Since the magnitude of β will (in general) vary with frequency, the feed-back circuit can exert an influence (which may be considerable) upon the overall frequency/response curve.

The presence of the feed-back connection across the output can also modify the output impedance in some cases.

Circuits for Negative Feed-Back. So many arrangements are possible that only one or two typical ones can be mentioned.

In radio-frequency amplifiers, the feed-back frequently consists simply of a small condenser between appropriate points (for example, between the anodes of two successive stages). A simple arrangement for a single-stage amplifier deriving its grid bias from a cathode resistor is shown in Fig. 271, where R_1 and R_2 form the bias resistor but only R_2 is shunted by the usual large condenser. The A.C. components of the anode current therefore produce voltages across R_1 (which are applied to the grid). C_1 reduces the negative feed-back at the higher frequencies and hence keeps up the response at frequencies for which it would fall due to the characteristic of the amplifier alone. For the same reason, L improves the response at the lower frequencies. Thus the application of negative feed-back in this case has greatly improved the frequency/response curve, has reduced distortion and made the amplification obtained much less dependent upon supply voltages.

Feed-back is sometimes applied to broadcast transmitters by rectifying a portion of the modulated radio-frequency output,

thus extracting the modulation, and applying this to the modulation input (in the correct phase to produce negative feed-back). Any distortion or "hum" introduced by the modulator and modulated amplifiers is thereby greatly reduced.

The Keying of Telegraph Transmitters. The particular problem of the keying of a telegraph transmitter for "mark" and "space" conditions is of course a special modulation problem. Since signal formation is of no real importance, but only a clear distinction has to be made between the "mark" and "space" conditions, we do not need to use "fidelity" modulation equipment for such a purpose, but simpler circuits are possible.

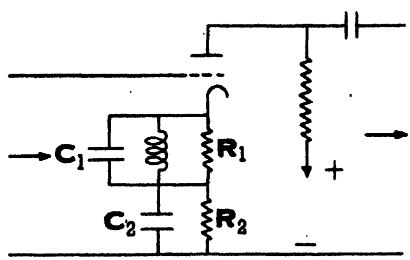


FIGURE 271.

The requirements for satisfactory keying of a telegraph transmitter may be summarised as follows: No change of frequency due to keying; no radiation on "space"; no large voltage surges, and no interference produced in neighbouring receivers tuned to a different frequency.

To satisfy the first requirement it is necessary to key only final or the higher power amplifier stages, and have the drive oscillator (also, preferably, the low power amplifier stages) running continuously.

Voltage surges will occur if the keying arrangements are such that the power drawn from the supply is very different during mark and space and if the power supply system cannot follow these changes with sufficient rapidity without surging.

Interfering noises, termed "key clicks," may be heard in neighbouring receivers even though these are tuned to quite a different frequency from that of the transmitter. This is because the starting and stopping of radiation at the beginning

and end of a mark is too sudden, so that the transient contains a great number of wide-spread frequency components. The remedy, therefore, is to adjust the keying system so that the growth and delay is rapid enough to give satisfactory signal formation at the distant receiver but not so rapid as to spread interference over a wide frequency-band.

For the keying of low-power transmitters at hand speeds, very simple arrangements are satisfactory. When an A.C. supply is used, for example, the Morse key may be placed in the primary circuit of the transformer supplying the receiver. Alternatively, the key may place a large bias on the grid of the amplifier valve.

When high-power transmitters are required to work at high speeds, more elaborate methods will be necessary, since the keying must be carried out by a light telegraph relay and surges and key-clicks will be much more serious. The keying arrangement adopted will be very dependent upon the type of power supply.

If a hard valve rectifying system is employed, for example, then the regulation will be poor (that is, there is a large change in D.C. supply voltage when the load changes). Also, due to the large smoothing system, the " electrical inertia " of the arrangement is large and it cannot respond quickly to sudden changes of load. In such cases it may be necessary for satisfactory high-speed keying, to provide an alternative (absorber) load during spacing so that the load on the supply is constant and transient surges are avoided.

The regulation of mercury-arc and hot-cathode, mercury-vapour rectifiers is considerably better and so is that of large D.C. generators.

When the supply permits, a simplified, partial-absorption circuit may be successfully employed.

Absorber Keyed Circuit. The connections of an absorber keying circuit are shown in Fig. 272.

In shunt with the main supply is a valve R_a, the absorber valve, in whose anode circuit is a resistance R. From the bottom end of the resistance R, namely " A," the feed to the No. 2 amplifier is taken.

On mark, the absorber valve is cut off by negative on its grid, and thus the only current flowing through R is the feed

to No. 2 amplifier. This will reduce the voltage to the anode of this stage, but if R is not too large, sufficient voltage can be obtained to give full output from this valve. If No. 2 amplifier is being driven from a previous stage, it will drive the No. 3 stage, the latter giving full H.F. output and in consequence taking full feed from the main supply.

On space, the absorber valve is made conductive by the application of positive to its grid. The effect of this is twofold. Firstly power is absorbed from the main supply, and secondly the feed, in passing through R, causes the voltage to No. 2 amplifier to fall, and if the value of R and absorber feed are so adjusted, the volts supplied to No. 2 amplifier can be reduced to almost zero. Now if the No. 2 amplifier has no

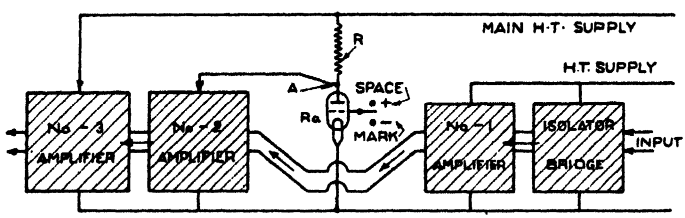


FIGURE 272.

feed current, although it is still being driven, it is unable to drive No. 3 amplifier, and in consequence the latter ceases to give H.F. output, and if the valves are backed off by static negative bias they will cease to draw feed from the supply.

Keying can be effected, then, by a simple change of voltage on the grid of the absorber valve, this voltage change performing the following operations:

Mark (G_b—)	*Space* (G_b+)
(1) Absorber circuit cut off.	(1) Absorber circuit takes feed.
(2) Volts supplied to No. 2 amplifier which drives No. 3 amplifier.	(2) No. 2 amplifier cut off.
(3) Amplifier No. 3 takes feed.	(3) No. 3 automatically cut off and takes no feed.

The correct adjustment of an absorber circuit is one which maintains constant voltage of supply on both "mark" and "space" and if the supply has a bad regulation this means obtaining exactly equal total feeds; whereas if the regulation of the supply is good, the space absorber feed need not be so large.

With such a system, although the power keyed at the absorber grid is small, it is by no means negligible on large power sets, and if really high speeds are desired a sub-absorber valve becomes necessary. This is virtually a stage of resistance-capacity amplification, the keying on the sub-absorber operating on the main absorber in opposite phase. Thus, for space, the sub-absorber valve is made positive and on mark, negative.

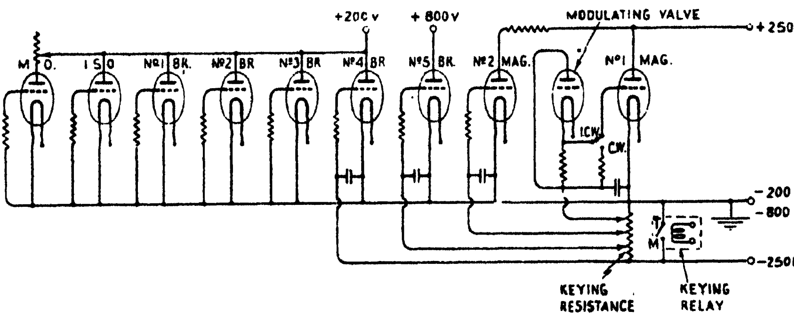


FIGURE 273.

This absorber system, whilst it appears a very elaborate method of keying, is fully justified by the results it has given, and transmitters of the greatest power used can be keyed at the highest speeds. The only key necessary is a light form of telegraph relay which can be operated by the incoming telegraph-line current direct, and as a matter of interest it may be recalled that during the Post Office acceptance tests of the Grimsby Beam Station transmitting to Australia, speeds of over 350 words per minute were maintained over long periods.

Partial Absorber Keying. This method utilised the main amplifier valves as absorbers during spacing, the feed then taken by them cutting off a preceding amplifier in the usual way.

One method of accomplishing this may be seen by referring to Fig. 273, from which it is observed that there is a resistance between cathode and negative H.T. of the last two amplifying

stages (marked " No. 2 mag." and " No. 1 mag."). The keying relay leaves this resistance in circuit during space but shorts it during mark. During space, therefore, the feed of these two amplifiers flows through the resistance, thereby providing self-bias but also providing sufficient grid bias to the stages marked " No. 4 *BR* " and " No. 5 *BR* " to cut them off. Hence the last two stages are not being driven, there is no output from the transmitter, and these stages take a static feed of value depending upon the tapping on the keying resistance.

Since the feed of any given valve operating under Class C conditions can be considerably greater than its static feed it is clear that a keying system of the type described can only absorb a small proportion of the total power. This means that unless the regulation of the supply is reasonably good, voltage fluctuations will occur in the set causing possible breakdowns and frequency scintilation.

Frequency Modulation. It has been explained that frequency modulation is sometimes adopted as an anti-fading device for telegraph signals. One method of applying frequency modulation is as shown in Fig. 274, this being the arrangement adopted on telegraph transmitters employing the Franklin master oscillator. *LC* represents the oscillatory circuit of the drive, to the inductance of which is coupled a plate C_2 having a very small capacity to the winding. This plate is connected to earth through a diode valve V_1 and it is clear that any variation of feed through the diode will vary the earth capacity effect of C_2 to the inductance, and this in turn will vary the frequency of the master oscillator. Fig. 275 gives an idea of the relationship between diode feed and frequency change, from which it is clear that if we wish equal changes in anode current to produce equal changes in frequency we must set the diode on the straight part of the curve (O, say), by supplying a D.C. potential. The steepness and straightness

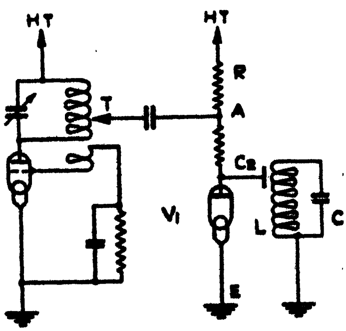


FIGURE 274.

of the curve connecting frequency shift and diode current depends largely upon the position of C_2 along the inductance.

The source of the modulation is a note oscillator of frequency f_s, the output from which is applied to the diode circuit (see Fig. 274) so that the diode current is varying at a frequency of f_s about the mean value. Hence the frequency of the master oscillator is made to vary between certain limits $f_c + f_n$ and $f_c - f_n$ the change occurring f times a second.

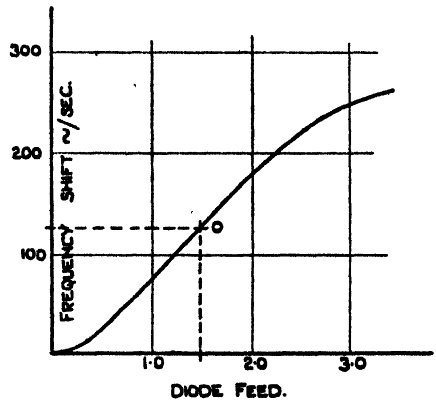


FIGURE 275.

The frequency change f_n will evidently depend upon the amplitude of the voltage applied from the note oscillator since this determines the amount of variation of diode current and the position of the tap T will, therefore, control the value of f_n.

It should be remarked that as the master oscillator frequency will be multiplied, any frequency variation impressed on it will also be multiplied the same amount, but the rate of variation (the modulation) will not be multiplied, and this must be allowed for in setting up the circuit. Generally speaking the frequency variation finally required is rather more than the modulation.

A telegraph transmitter employing frequency modulation will be keyed in the normal way in a stage subsequent to the master oscillator (usually by the absorber method) so that the transmitter is still working on the principle " full radiation on mark "—" no radiation on space," but the radiation on " mark " takes place on a small band of frequencies.

SELECTED REFERENCES.

1. *Series Modulation.* Ditcham. Mar. Rev. No. 41.
2. *The Marconi-E.M.I. Television System.* Blumlein, Browne. Davis and Green. *J.I.E.E.*, Vol. 83 ; 1938.

3. *The Empire Service Broadcasting Station Daventry.* L. W. Hayes, B. N. MacLarty. *J.I.E.E.*, Vol. 85; 1939.
4. 100 *kW Short Wave Broadcast Transmitter.* Green and Moody. Mar. Rev. No. 74.
5. *Frequency Modulation Equipment.* Green and Hewitt. Mar. Rev. No. 57.
6. *Cathode Modulation.* Jones and Edmonds. *Q.S.T.* Nov., 1939.
7. *An Economic Feed-Back Circuit.* Sandy and Twatt. Mar. Rev. No. 72.
8. *Properties of Negative Feed-back.* Farren. *W.E.* January, 1938.
9. *Negative Feed-back.* Chaffee. *B.S:T.J.* 1939. Page 404.
10. *Relations Between Attenuation and Phase in Negative Feed-back Amplifier Design.* Bode. *B.S.T.J.* 1940. Page 421.

CHAPTER XIV

PROBLEMS OF RECEPTION

THE various classes of traffic (as indicated below), which may have to be handled by a receiver, show the variety and complexity of the general problems of design.

(*a*) Broadcasting.
(*b*) General purpose telegraph and telephone reception by a skilled operator.
(*c*) High speed telegraph signals for relaying and recording at a Central Telegraph Office.
(*d*) Telephone signals suitable for linking together the public telephone networks of various countries.
(*e*) Television.

Generally speaking receivers evolved for duty in any one of the above mentioned groups will only be suitable for that particular purpose and it is not an economic proposition to develop a universal type of receiver.

In all receivers, the following are desirable features: Freedom from noise, selectivity, fidelity, sensitivity, and ease of control. How near the receiver approaches the desired ideal will depend upon the purpose for which the receiver is to be used and upon any economic considerations which may be imposed. For instance receivers in (*a*) and (*e*) are for entertainment purposes and must have high fidelity, a very simple system of control, are often required to work with any aerial, but they need have only low sensitivity as they will work from high field strength levels and probably in noisy localities. Receivers in (*b*) will have to be capable of quick searching, and even in the present state of the art will often have to work with simple transmitters which are not too constant in frequency. Sensitivity may have to be high or low and as both telegraphy and telephony may be used, methods of controlling selectivity

are desirable, and a quick change from a limiting to an automatic volume control is necessary. Evidently the requirements of (*c*) and (*d*) are very exacting and the nature of the services so important that elaborations are here admissible which would be out of place in the other groups. Some further discussion of receivers suitable for (*b*), (*c*) and (*d*) will be found in Chapters XV and XVI, although we shall deal here with the principles on which they are built. The peculiar problems of (*e*) are in providing the very large band-width necessary and in keeping down phase distortion.

Although we assume the reader to be familiar with the ordinary principles of reception we propose to give a preliminary survey of the problem, describe the outline features of the three prominent types of receivers used, namely "Straight," "Super-regenerative," and "Superheterodyne"; and then to call attention to features which are important in short wave and ultra-short wave work.

Noise. Noise may be produced by causes which are external to the receiver or internal. Atmospheric interference with reception has already been discussed (page 119) and it has been pointed out that interference is caused on all frequencies, though the amplitude of the higher-frequency components coming within the short-wave band is much smaller than those in the long. Since all frequencies are present, it follows that the wider the band-width of the receiver the more energy will it pick up from the atmospheric disturbance and thus it is essential to design the circuits such that they do not employ a greater band-width than necessary. External noise in the form of "man-made" interference has also to be considered in most situations and is worse the shorter the wavelength. All arrangements which produce sudden changes of current, especially if the change is accompanied by sparking, may produce the radiation of electromagnetic waves. Such waves will be heavily damped and may therefore be considered as a wide band of frequencies, the mean frequency (or frequencies) being determined by the effective inductance and capacity associated with the connecting leads to the apparatus producing the disturbance. Interference due to sparking brushes in a motor or similar causes can usually be reduced to inoffensive proportions at the source by the fitting of a

condenser, or a condenser-resistance unit across the motor terminals of sufficiently low H.F. reactance to make the production of an H.F. voltage practically impossible. The B.S.I. have recently suggested standard methods for measuring the interference produced by small motors and propose to issue a distinguishing mark to electrical appliances the interference from which is below a certain value.

When a receiver is supplied from the mains, low and high frequencies may be brought into the receiver by way of the mains and the inclusion of filter circuits in the supply may be necessary. In cases of bad sites it may be desirable or even necessary to move the aerial to a favourable position and connect it to the receiver through a screened feeder, the improved signal/noise ratio outweighing the attenuation introduced by the feeder.

A powerful source of interference is that due to motor car and aircraft ignition equipment. This has the elements of a heavily-damped, spark transmitter and the electrical constants of the high-tension leads employed makes the interference most severe on wavelengths of about 6-7 metres. When wishing to use a receiver on any frequency in a car or aircraft, some form of suppression will be necessary. On aircraft engines this problem has been solved by the complete screening of all ignition leads right up to the engines and including the sparkling plugs and the bonding of the so-called screening harness according to a definite plan.

Internal Noise.[3] If all sources of external noise could be eliminated, there would still be noise generated in the receiver itself and thereby setting a limit to the weakness of incoming signal which can be handled if a satisfactory signal/noise ratio is to be obtained at the output. Assuming spurious noises due to faulty components, valves, batteries, etc., to be eliminated there will still remain what is known as circuit-fluctuation noise. This includes

(*a*) Thermal-agitation noise.

(*b*) Shot effect and flicker effect.

Thermal-agitation noise is due to the movements of the electrons in the material of which the circuit is constructed. This results in the production of a very small voltage across

the circuit, having components at all frequencies. The magnitude of the thermal-agitation voltage across an impedance is proportional to the resistance component of the impedance. While this effect takes place in all the circuits, the only important case is that of the tuned circuit preceding the amplifying stages in a high-gain receiver. In the case of a superheterodyne receiver, if the initial radio-frequency, or signal-frequency as it is often called, tuned circuit is brought into tune with the rest of the receiver a pronounced increase in output noise should be heard, this being a measure of the sensitiveness of the receiver, and also of the extent to which other sources of noise have been eliminated. The fluctuation voltage produced across a given circuit remains constant with time, and is used in some measuring equipment as a means for adjusting the apparatus to give a standard amplification.

Shot effect is the noise introduced by the valve, due to the random arrival of electrons from the cathode which means that the anode current is not continuous. As with thermal-agitation noise, shot effect is only important in the first valve.

Selectivity. The curve connecting output with frequency in an ideal receiver would be rectangular, that is, all the frequencies comprised in the wanted signal would be passed through the receiver with relative amplitudes unchanged and all unwanted frequencies would be rejected entirely.

Another way of expressing this would be to say that the energy level is constant throughout the pass-band and that for all frequencies outside this the energy level is down by an infinite number of decibels.

This ideal cannot, of course, be attained and an actual curve might be as Fig. 276 (produced by a pair of over coupled circuits). The selectivity may be specified in the following manner: State the pass-band (f_p) and the variation of level that may be permitted (a_p db.). Specify the drop in level (a_e db.) at a frequency (f_e) slightly outside the pass-band. The greater the difference between a_e and a_p and the less the difference between f_e and f_p the better will be the circuit characteristic. For example, in a commercial telephone receiver the values of the variables would be of the order of $f_p = 5$ kc/s. $a_p = -6$ db. $f_e = 7{\cdot}5$ kc/s. $a_e = -80$ db. overall.

In the case of commercial receivers of high grade, such selectivity would be obtained by groups of over-coupled tuned circuits, but in the simpler class of receiver we would have to be content with a less elaborate circuit giving an inferior type of selectivity as shown dotted in Fig. 276.

Fidelity. As was mentioned in Chapter I, the ideal receiver should impart no kind of distortion, i.e. amplitude,

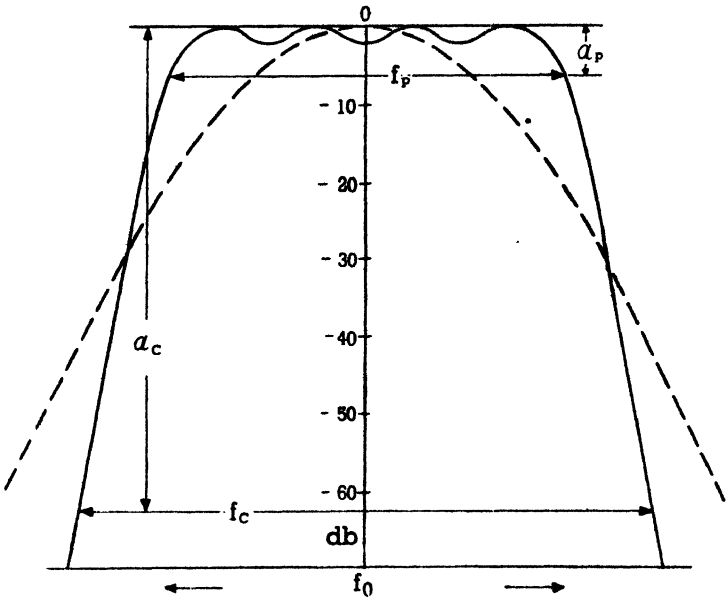


FIGURE 276.

phase, or harmonic. In audio-frequency work, phase distortion is of no consequence but it must not be allowed in television; a certain amount of amplitude distortion can be permitted as the ear and eye is not very sensitive to change of amplitude, but harmonic distortion must always be avoided, as this generates new frequencies which are quickly noticed. Low harmonic distortion entails careful design of the detector stage of the receiver, and the audio-frequency circuits. It will be appreciated that the fidelity of a broadcast receiver must be high and, therefore, not more variation of level than 2 db. can usually be permitted throughout the pass region; in receivers

for commercial telephony variations of level of 6 db are usually permitted.

Sensitivity. Receivers designed for different duties will differ widely in the sensitivity. Broadcasting stations on the normal broadcast wavelengths provide a very high level of field strength within their service areas, and because broadcast receivers often work in locations where there is much electrical interference, high sensitivity should not be necessary and may, in fact, be a disadvantage as it encourages the use of very inefficient aerials with a resulting poor signal/noise ratio.

In the case of receivers for commercial services, because of the reduction of power of commercial transmitters to the barest minimum, the design calls for the highest possible receiver sensitivity, necessitating a careful choice of site, and circuits with a low internal noise level.

Ease of Control. Apart from a commercial receiver, where the same station will probably be worked for hours at a stretch and where the high overall performance justifies some elaboration of manual controls, the average receiver should have as few manual controls as possible, particularly so in the case of a broadcast receiver. Where automatic controls replace manual they should give results at least equal to those which can be achieved by a skilled operator. Automatic gain control, that is, the control of the overall gain of a telephone receiver so as to keep the output constant when the input varies, is greatly superior to any manual control of volume. Other forms of automatic control which are coming into wider use, are: automatic frequency control which aims at keeping the receiver in tune; and automatic selectivity control which is designed to keep the pass-band to the minimum width necessary for the signal being received.

We will now discuss the three types of receivers previously mentioned.

The Tuned Radio-Frequency Receiver usually called "Straight." This type of receiver has one particular advantage, it cannot create interfering signals, and whistles in the output can only be created by two carriers coming within the audio frequency range. It has a number of disadvantages however:

(1) Each tuned circuit needs its own variable condenser, and difficulties of ganging limit the number of stages possible where simplicity of control is wanted.

(2) The amplification of any stage is dependent on the resonant-frequency impedance ($Z = L/CR$) of the tuned circuit in the anode of each valve. Owing to the large variation of C over the tuning range the resonant impedance (and amplification) vary considerably. Thus sensitivity is low at the low frequency end of the range increasing to the high frequency.

(3) Attenuation at a given off-tune frequency is proportional to Q/f. Since Q changes but little, attenuation is roughly inversely proportional to frequency, and the selectivity falls as the carrier frequency rises.

Poor sensitivity and selectivity can to some extent be corrected by the use of reaction, usually at the detector stage, but since a selectivity curve with a sharp peak is produced thereby, it is then unsuitable for broadcast reception, but may be suitable for general purpose work in the hands of a skilled operator.

This type of receiver has a certain vogue, is used by the amateur experimenter for quick searching and the type has also been used for television reception. Its success in this field is largely due to the fact that the tuning frequency is fixed, since only one station is to be received, a wide pass band is required for the vision, and interference due to oscillator harmonics (such as are produced in the super-heterodyne) completely absent. Otherwise it is not a good receiver for short-wave work because both selectivity and sensitivity are low. Sensitivity is low because this depends on the impedance of the anode tuned circuits; and because of the high minimum capacitance of the circuit, the tuning coil on short waves is very small, This means, of course, a low anode impedance, of the order of only 20,000 ohms on 30 metres say, as against over 100,000 ohms over the medium wave band. Apart from this, sensitivity is limited by the low input-impedance of the valve, which may be comparable with that of the circuit. This is mainly due to feed-back of voltage from the cathode lead inductance into the grid circuit through the grid-cathode

capacitance, the feed-back voltage being in phase opposition to the input voltage. As one approaches the ultra-short wave band electron transit time causes a lag in the control of grid voltage on the electron stream, and this lowers the grid/cathode resistance and therefore adds to the damping of the tuned circuits.

Selectivity is poor as well, also due to the unfavourable L/C ratio of the tuned-anode circuits used; thus 10 kc/s off tune may give a ratio of 5×10^{-4} at 20 Mc/s, as compared with only 10^{-2} at 1 Mc/s.

A typical circuit for a "straight," battery-operated receiver is shown in Fig. 277. The radio-frequency, tetrode stage usually

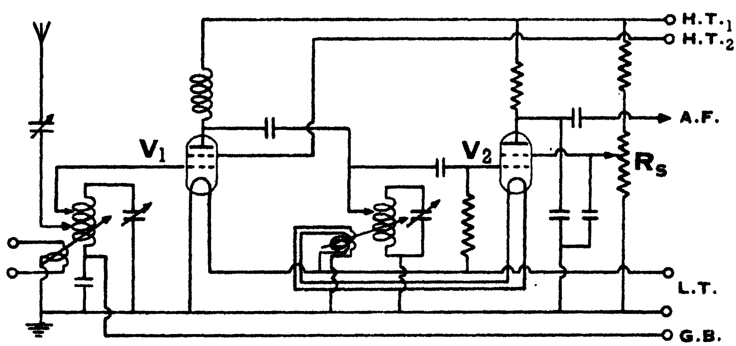


FIGURE 277.

has but little amplification, and serves mainly as an isolator to prevent reflection of aerial impedance variations into the detector tuned circuit, and to reduce re-radiation when the detector circuit oscillates. Both aerial and valve are tapped down low on the first tuned circuit so as to improve selectivity, and the aerial being of low impedance is usually tapped at a lower point than the valve. Also since the aerial is not tuned, its impedance will vary over the wave-range, and by tapping it low down, this variation does not influence the rest of the circuit to any extent. The detector valve is of the tetrode type because of its high gain, and its low anode-grid capacitance reduces anode damping on the grid tuned circuit. For the normal short wave receiver, the variable condensers will have a maximum value of only about 100 $\mu\mu F$ as, although such small

condensers restrict the tuning range for a given coil, it enables the impedance to be kept high and reduces the variation of sensitivity and selectivity over the wave-range, this being covered by groups of coils.

Smooth control of reaction is difficult to obtain on short waves but is very essential to a successful receiver of this type, since it will be most sensitive for telephony reception when it is just free from oscillation, and for C. W. telegraphy when it is just oscillating. In the circuit shown, the detector grid leak is returned, as usual, to the positive side of the filament battery and reaction is provided by parallel feed-back coils in both filament leads, the reaction being varied by adjustment of the screen-grid potentiometer R_s.

Super-Regenerative Receiver.[5] Suppose an oscillating detector to be adjusted so that it is on the verge of oscillation. A small signal voltage applied to the grid will now "shock" the unstable arrangement, and oscillations will commence which will build up to the maximum amplitude permitted by the valve constants. It follows that the anode current has been changed considerably by a small change of grid volts—in other words, the arrangement is a very sensitive type of "trigger" relay.

Although the final amplitude to which the oscillation would grow is determined only by the valve, the rate of rise is proportional to the amplitude of the input signal and inversely proportional to the value of the tuning inductance.

Such an arrangement would not be capable of reproducing a signal, however, because the oscillations once started would not cease at the conclusion of the signal, since the effective resistance of the circuit is negative.

Armstrong showed, however, that if the resistance of the circuit could be varied periodically (at a frequency considerably lower than the signal frequency) between positive and negative values, then very large amplification could be produced without distortion of an incoming modulated signal. Fluctuation noise will also tend to build up in such a circuit with the result that a hissing sound will always be heard in the telephones.

It will be evident that if telephony is being received, the quenching frequency must be above the audible limit. The

greater the difference between the quenching frequency and the signal frequency, the greater is the amplification, because the signal has a greater time to build up to a greater value during the half cycle of the quenching which makes the resistance of the circuit negative. The super-regenerative principle is therefore more suitable for high than low frequency reception, and in fact is almost confined to ultra high-frequency work, for which compact, sensitive receivers can be designed.

The super-regenerative receiver has a high noise level and is therefore only suitable for "commercial" quality reproduction. It is un-selective, and it is not possible to insert tuned circuits before the oscillator in order to increase selectivity, If this is done, it is found that the tuned circuit has large oscillations set up in it when the super-regenerative circuit has negative resistance, but these oscillations carry on when the quenching is applied, and set the super-regenerative circuit oscillating again when the quenching is removed, irrespective of whether the original signal has ceased or not. Of course, this effect could be prevented if a perfectly "non-reversible" coupling could be devised between the tuned circuit and the super-regenerative circuit.

At the present stage of ultra-short wave development, however, this lack of selectivity is advantageous, as it facilitates searching for a station in the enormous (and not at all crowded) frequency band involved.

Fig. 278 shows the circuit diagram for a suitable super-regenerative receiver for ultra-short wave work. The aerial is connected through a screened feeder to the closed LC circuit by means of a variable mutual coupling, the adjustment of the latter providing a sensitive control of signal strength. The triode valve V_1 is the super-regenerative detector, and the triode valve V_2 the quenching valve. The grid leak of the detector is returned to the positive side of the filament, and the "Hartley" type oscillator is controlled by the condenser C and the detector anode voltage (by R). This type of circuit is simple and free from spurious resonances which may cause squeggers. For the best results and to keep the noise level to a minimum, the coil should have a high Q value so that minimum reaction is necessary. The quenching voltage is

injected into the anode circuit rather than the grid as it has been found this gives more stable operation, the quenching circuit being an ordinary back-coupled oscillator at a frequency of 20 kc/s, a value low enough not to reduce sensitivity and cause whistles, but high enough to allow of easy filtering in the audio-frequency stage.

The Super-Heterodyne Receiver.[1] This system has the advantage that nearly all the amplification may be provided at a fixed frequency, thereby reducing adjustments and simplifying the design of the amplifying circuits. The band-width of the receiver may also be kept the same over the whole

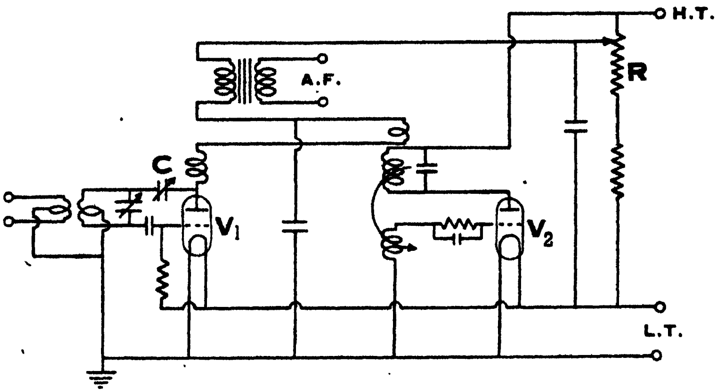


FIGURE 278.

wave-range for which the receiver is designed, this being difficult in the "straight" receiver.

In the super-heterodyne, the desired carrier frequency, together with a local-oscillator frequency, is applied to a frequency-changer valve which produces in its anode circuit components of the sum and difference frequencies of the two oscillations. Either of these frequencies, which carry the modulation, may be selected by suitable filters, although the difference-frequency is always chosen, and amplification and selection after the frequency-changer will thus be carried out at a fixed frequency known as the intermediate frequency (I.F).

A consideration of Fig. 279 will show that the same selectivity can be imparted to a receiver by one or two stages tuned to

the I.F. as by many tuned to the radio frequency (R.F.). Curves (*a*) and (*b*) are for two circuits having equal damping but tuned to R.F. of 10 Mc/s and an I.F. of 200 kc/s respectively. When the curves are combined it is seen the overall selectivity is very high.

Sufficient tuning at the R.F. must be provided, of course, to ensure that the "image frequency" or "second channel" interference is cut out; that is, an unwanted frequency having a frequency differing from that of the wanted carrier by twice the I.F. must be rejected before passing to the frequency-

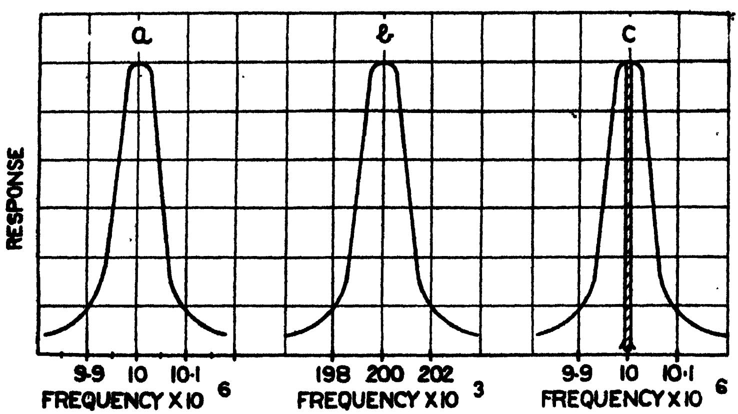


FIGURE 279.

changer, which would be unable to differentiate between it and a wanted signal.

The choice of I.F. to use in a receiver is governed by the following considerations. A low I.F. makes it easier to produce a very selective receiver but more R.F. tuning is required to eliminate the image frequency, whilst a very high I.F. will make the provision of the necessary amplification more difficult. The use of a low I.F. makes it more difficult to prevent interaction between the beating oscillator and the signal tuned circuit, and for a general purpose circuit an I.F. of about 450 k/cs is usual.

In a frequency-changing system interference can be caused by the interaction of the harmonics of the frequencies beating together, owing to the fact that the frequency-changing valve

must have asymmetrical characteristics and therefore produces harmonic distortion. Such interaction (which may result in audible whistles) is reduced by the first circuit tuning, by a careful choice of the I.F. and by not allowing the oscillator amplitude to be too large.

The super-heterodyne lends itself to automatic control of volume, tuning, and selectivity. It is, therefore, in many ways superior to the other types of receiver, its only undesirable feature being that just mentioned, namely, its ability to produce unwanted audio frequencies through the frequency-changing system.

A description of a super-heterodyne suitable for commercial work is given in Chapter XV, but in Fig. 280 is shown the

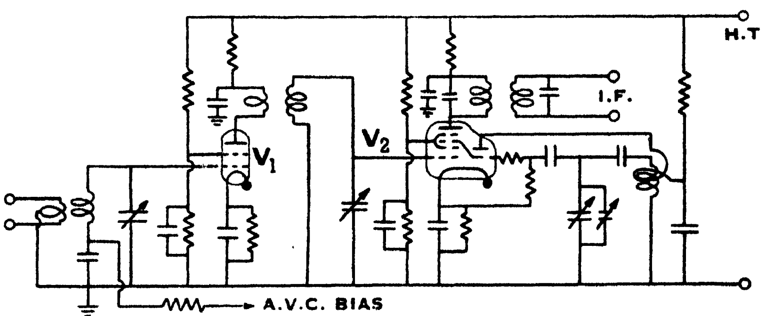


FIGURE 280.

radio-frequency and frequency-changing circuit of a simple type suitable for a general purpose receiver. One stage only of radio-frequency is shown using an aligned screen-grid valve, and the frequency-changing valve is a triode-hexode. One of the difficulties in designing a super-heterodyne for short waves is that at high radio frequencies, the ratio between oscillator and incoming frequencies decreases, and interaction between the two circuits becomes greater. This may be due to inter-electrode capacity or electron stream coupling. This has two effects; it causes the R.F. circuit to reflect a resistance and reactance component across the oscillator circuit, the first affecting oscillator amplitude, the second oscillator tuning; it also causes an oscillator voltage component to appear across the R.F. circuit, and as this is usually out of phase with the oscillator electrode voltage it reduces the effective oscillator

voltage applied to the frequency changer. In the heptode frequency-changing valve, where the oscillator electrode is nearest the cathode, electron coupling has the same kind of effect but it is opposite in direction to that exercised by inter-electrode capacitance. At one particular frequency and set of operating conditions electron coupling may be completely cancelled by capacitance coupling and a condenser of 1 $\mu\mu F$ is sometimes connected externally between signal-frequency and oscillator electrodes to reduce this coupling effect over a range. In the hexode type of valve, this type of electron coupling is almost absent because the control grid is nearest the cathode, but there is another effect which must be allowed for. Here the oscillator voltage repels some of the electrons with sufficient velocity to be collected by the control grid when its voltage has the least negative value. The frequency for which favourable conditions for the collection of electrons arises increases with increase of signal frequency, so that a bias of about $-2{\cdot}5$ volts may be necessary at 20 Mc/s to prevent grid current whereas only about $-0{\cdot}5$ volt is necessary at 1 Mc/s.

The selectivity of any radio-frequency tuned circuit following the first valve should be high, not only to remove the image signal but to prevent noise side-bands and undesired signals beyond audio range of the desired carrier from reaching the frequency-changer, where they can combine with the oscillator or its harmonics to produce interference frequencies.

For the oscillator either a "Hartley" circuit or a tuned-grid is used. Such circuits are easier to maintain in oscillation over a range of frequencies than a tuned anode circuit and they give constant output. In the tuned-grid circuit tight coupling between feed-back and tuning coil is essential in order that the feed-back inductance should not approach the inductance of the main tuning coil, as if this occurs squeggering and blind spots result at high frequencies due to resonance of the feed-back coil and stray capacitance. It is usual to interleave the turns of the feed-back and tuning coils. The tendency to squegger can be reduced by using a low value of grid leak (not greater than 50,000 ohms) and grid condenser (not greater than 50 $\mu\mu F$) and a resistance of 50 ohms is often included in series with the oscillator-valve grid to reduce oscillation amplitude at high frequencies. The oscillator section of the

valve should have a high mutual conductance so that feed-back may be reduced to a minimum. It is because the coupling of oscillator to signal frequency needs to be weak that a separate valve is often used for the oscillator instead of the combined frequency-changer.

On short waves image rejection becomes a difficult problem, and can only properly be solved by increasing the number of R.F. tuned circuits, or by raising the I.F., or by employing the double super-heterodyne principle. In the last case a high I.F. is first employed, followed by a second frequency-changer and a lower I.F., but (except on ultra short waves), this arrangement is only usually employed for telegraph receivers.

In the case of short wave broadcast reception, the desired radio frequencies are limited to narrow frequency bands at specified points in the short wave range, and some receivers developed for broadcast purposes have what is known as band-spread tuning, which as its name implies, is designed to spread each of these narrow bands over the tuning capacitance range. This is achieved by the use of padding condensers which restrict the tuning range of the variable condensers.

A Special Type of Super-heterodyne Receiver. A super-heterodyne type of receiver employing an intermediate frequency higher than the signal frequency has been developed under various names, including "Infradyne" and "Single-Span." The object is to reduce, or eliminate altogether, the necessity for adjustable tuning at the signal frequency by making the image frequency very different from the required signal frequency. The method is difficult to apply to short waves because of the very high frequency oscillator and intermediate frequency amplifier required, but has been employed by the British Post Office in the construction of a "Quick Search" receiver for short wave, telegraph traffic from ships. Rapid searching over a very wide frequency band is essential for this service, and yet considerable selectivity is required when taking traffic.

Suppose the receiver to be required to receive up to 25 megacycles, the first intermediate frequency amplifier might work at 30 megacycles, the first oscillator then having a range of 30 to 55 megacycles. The image frequency is now 60 megacycles removed from the signal frequency, and hence the

input circuit may be a single, highly-damped, resonant circuit to provide a reasonably efficient input arrangement without critical tuning adjustment. The amplification at 30 megacycles will, of course, be small and it will therefore be necessary to change the frequency again to a much lower value and provide further amplification at this frequency.

Ultra-Short Wave Receiver for Television. When designing a super-heterodyne receiver for television on U-S.W. certain special features must be noted. Television transmission in England is carried out at two U-S.W. carrier frequencies, the vision at 45 Mc/s and the sound at 41·5 Mc/s. For the proper reception of these by super-heterodyne great care must be exercised in the selection of the intermediate frequency as vision and sound inter-action and I.F. harmonic interference is liable to superimpose patterns on the television screen. The minimum value of I.F. is determined mainly by the frequency separation between sound and vision carriers and it must be outside the range of possible direct interference due to inter-action between the two transmissions at the frequency changer. If the two carriers only are considered the I.F. must exceed 3·5 Mc/s and if we take into account the vision sidebands extending $\pm$ 2 Mc/s on either side of the carrier, it must be greater than 5·5 Mc/s. Harmonics of the I.F. may be produced at the detector and fed back to the R.F. stages. The I.F. must therefore be chosen so that inter-action of its harmonics with the oscillator at the frequency changer is outside the range of the I.F. amplifier. For example, if we have an I.F. of 6 Mc/s, the possible interfering harmonics are the 7th and 8th, namely 42 and 48 Mc/s, but these inter-act with the oscillator to produce I.F.'s of 2 and 3 Mc/s, both of which are sufficiently far outside the pass-band of the I.F. amplifier. There is a choice of about four possible I.F. frequencies, 6·9, 8·2, 10, and 13, where interference is likely to be least.

The I.F. stages may consist of single resonant circuits tuned to the same frequency and damped to give the required band width, or pairs of over-coupled, damped, tuned circuits, or of single damped, resonant circuits staggered in tuning. Maximum stage gain, as has been explained, is obtained with the maximum L/C ratio, and this is determined by the stray

capacitance of the circuit components, On these ultra-short waves inductance trimming will be preferred to capacitance in order to keep up the highest possible L/C ratio.

Pairs of over-coupled circuits give the better band-pass effect, and although there is a loss of amplification as compared with single tuned circuits of similar constants, this is offset by the fact that the stray capacitance is split between primary and secondary so that a higher L/C ratio is actually possible. Furthermore, owing to the double-humped response due to the over-coupling, less damping is required for a given pass band. These factors tend to make the over-coupled circuit superior to damped, single circuits, although it is not possible to use inductance trimming for the former as this alters the mutual coupling.

Much the same result may be achieved with single circuits staggered and there is little to choose between these two methods.

In order to reduce the pass-band and hence make it easier to obtain sufficient amplification some television receivers pass only one side-band. In ultra-high frequency work it is more usual to operate the oscillator of the frequency changer at a higher frequency than the signal. Although a lower frequency tends to give more efficient conditions as explained previously, this means that the image frequency occurs at a lower frequency and in a range where interfering transmissions might be expected.

The detector circuit usually comprises a diode valve with a load resistance and bypass condenser modified to pass the high vision frequencies without serious attenuation. The values of these two components are of the order of 5,000 ohms and 20 $\mu\mu$F. Leads to the vision frequency amplifier and to the detector circuit must be carefully de-coupled and screened to reduce feed-back of the I.F. harmonics to the R.F. circuits.

Following the detector is usually a resistance-capacity coupled amplifier, and these stages usually require small series inductances in their anode circuits to neutralise the stray capacitance and preserve a reasonable level response up to 2 Mc/s. Some form of D.C. restorer, often a diode detector, is necessary so as to reproduce any change in the general level

of illumination of the picture before the signal is used to modulate the cathode-ray tube.

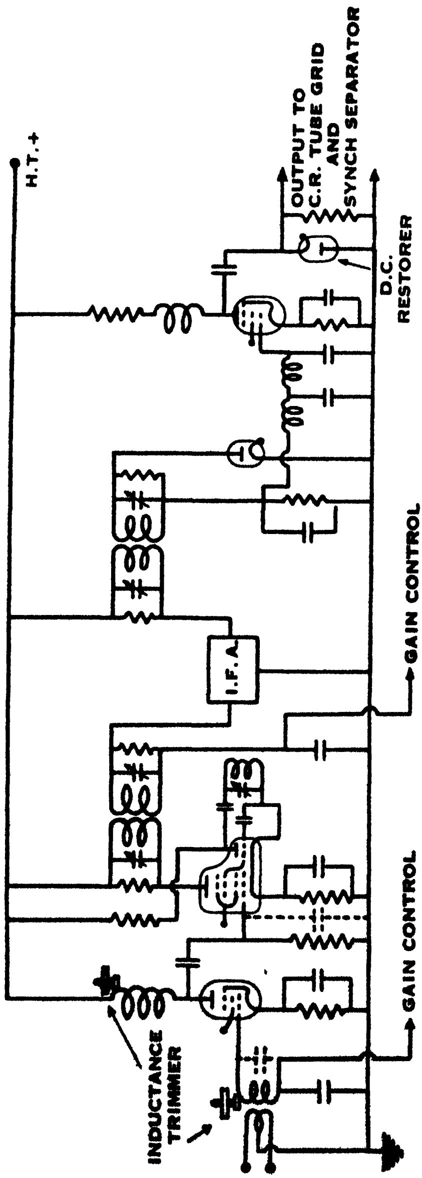


FIGURE 281.

The circuit diagram of a typical super-heterodyne circuit suitable for television is shown in Fig. 281.

General Considerations. As the carrier frequency increases, effects which are negligible at low carrier frequencies become important.

Stray capacitance and inductance may lead to loss or gain of sensitivity, detuning, or instability and, therefore, great care must be taken to make all connecting leads as short as possible. Earth connections to a common busbar rather than to indiscriminate points on a chassis, very efficient screening between components is essential, and nothing in the way of a loose contact or joint can be tolerated. Although it would be expected that de-coupling would be comparatively easy at high frequencies, since even a small capacity will have very low reactance, it has been found that very great care is necessary in the selection of condensers for this purpose. For instance the average paper

condenser is of either the rolled tinfoil type or a rolled paper strip sprayed on one side with a conducting material, usually graphite. The latter is quite useless for the purpose of shunting high frequencies, and the former is only useful if specially manufactured with a number of connections brought out along the rolled strip. At very high frequencies the inductive reactance of the connecting leads and plates may be greater than the capacitive reactance of the plates, so that the net effect across the condenser terminals is inductive.

A series of measurements by Dr. Hartshorn[4] at the N.P.L. showed that at 50 Mc/s, a 0·0005 μF mica condenser had a lower impedance than a 0·1 μF so-called "non-inductive" tubular condenser although in both cases the condenser was inductive. Smaller condensers (100 $\mu\mu$F) showed capacity reactance, the apparent capacity increasing with frequency. Condensers made with a dielectric using a ceramic material of large dielectric constant can be made in such small dimensions that inductive effects are negligible, and such condensers are particularly suited to high-frequency work.

It is interesting to note, however, that ordinary metallised resistances are still non-inductive at 100 Mc/s and have almost the same resistance value as at low frequencies.

Valve defects are much accentuated. Thus lead-inductance, inter-electrode capacities, electron transit time, all combine to produce a high grid input-conductance and phase shifts, causing loss of amplification and selectivity. For instance, tests on a conventional pentode valve showed that its input impedance fell from 3·3 megohms at 1·3 Mc/s to 0·0086 megohms at 50·4 Mc/s. This means that valves of conventional design can only be used satisfactorily down to about 6 metres, and then only if the grid is brought out to a top cap. For amplification below these wave lengths, the so-called "Acorn" type of design is much better, as this has closely spaced electrodes, thus keeping the usual relationship of r_a, μ, and g_m, but having well-spaced and very short connections through the glass seals as shown in Appendix V.

Receivers for Centimeter Waves. At frequencies corresponding to wavelengths of only a few centimeters, although amplification is not so far possible, valves and crystals may be

used as detecting devices, but every effort has to be made to obtain the greatest sensitivity with them.

It has long been known that a circuit precisely similar to the electron oscillator used for transmitting may also be used for reception, the output being taken from the anode circuit. The functioning of the receiver has not received nearly so much attention as the oscillator, and it has been usually assumed that the signal applied to the grid affected the electron oscillation and in some way this produced a detecting action.

Carrara[6] found, however, that the adjustments for best detection were, in fact, different from those for oscillation, and that the detecting action was apparently independent of electro

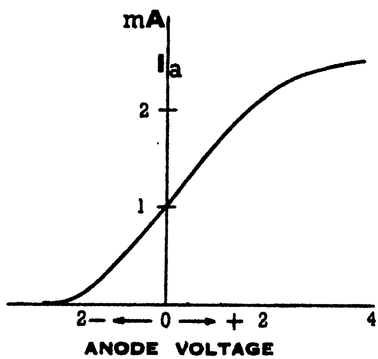


Figure 282.

transit time because comparatively low frequencies could be detected using the same values of valve voltages. He considered that, because so many electrons were brought to a standstill very close to the anode, there was a virtual cathode there. The equivalent of diode rectification takes place between this virtual cathode and the anode, and the arrangement is effective at extremely high frequencies because the distance between cathode and anode is very small and the transit time does not exert a harmful influence.

Hollman[7] examined the static characteristic of a valve with grid positive and anode negative and found it to be shown in Fig. 282. Evidently the bottom or top bends will produce rectification of an alternating voltage and it is therefore confirmed that the arrangement, termed by him a retarding-

field detector, can be used at any frequency. Hollman found that putting the output circuit in the grid circuit instead of the anode was a considerable improvement because impedance

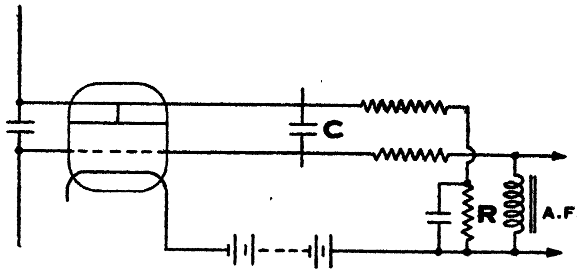


FIGURE 283.

in the anode greatly reduces the curvature of the dynamic characteristic compared with the static, thereby reducing the rectifying action.

When the applied voltage has a very high frequency, we can also arrange that the valve is on the verge of electron oscillation and hence there will be a reaction effect and the arrangement forms a very sensitive detector of centimeter waves. A circuit diagram of a suitable arrangement is shown in Fig. 283. The condenser C is in the bridge of a Lecher-wire tuning circuit, and the high resistance R keeps the anode voltage at a suitable value.

It has already been mentioned that the best adjustments for detection are not the best for oscillation. In order to improve performance, Hollman devised a two-valve arrangement, one valve acting as a detector, and the other as a reaction device.

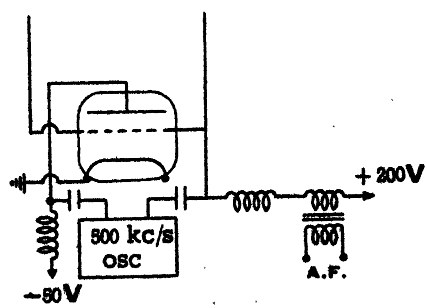


FIGURE 284.

The super-regenerative principle may be employed with the positive grid detector, the adjustments being such that an electron oscillation is produced but periodically quenched, Fig. 284, showing a resonant grid type of oscillator [9] employed in this way.

SELECTED REFERENCES.

1. *The Super-Heterodyne.* Witts. Pitman.
2. *Electric Circuits and Wave Filters.* Starr. Pitman.
3. *Spontaneous Fluctuation Voltages.* Moullin. Oxford.
4. *Properties of Chokes, etc., at Very High Frequencies.* Hartshorn and Ward. *J.Sc.I.*, April, 1937.
5. *The Super-Regenerative Receiver.* Scroggie. W.E. November, 1936.
6. *Detection of Micro-Waves.* Carrara. *P.I.R.E.* Vol. 24; October, 1936.
7. *The Retarding Field Detector.* Hollman. *P.I.R.E.* Vol. 22; May, 1934.
8. *Causes for Increase of Admittance, etc.* Strutt and Van der Ziel. *P.I.R.E.* Vol. 26; August, 1938.
9. *Micro-Ray Communication.* McPherson and Ullrich. *J.I.E.E.* Vol. 78; 1936.

CHAPTER XV

COMMERCIAL RECEIVERS

THE much lower level of atmospheric noise on short waves means that a satisfactory service can be carried out with a smaller field strength at the receiver, but because the received field strength is lower, greater amplification becomes necessary. The principal characteristic of the short-wave signal is the fluctuation of strength to which it is subject, due to reasons that have been discussed in Chapter V; and as far as possible the minimum value must give a sufficient output from the receiver. If occasional very "deep" but brief fades result in a not perfectly recorded letter on the telegraph tape, or an inaudible syllable in the subscriber's telephone, it is not of the greatest importance; but the effective amplification provided must be such that only these deepest fades cause trouble.

We shall expect, then, that the S.W. receiver for important point to point services will be provided with more amplification than a corresponding long-wave receiver, and the principal problems will be connected with providing this amplification, whilst increasing the noise level as little as possible. In connection with this it is of interest to note that a high-gain, modern commercial receiver has an internal noise level so low that it can give full output, of the order of 10 dbs. above 1 milliwatt, with a relative level of speech above noise of 10 db., when the field strength falls as low as one-tenth of a microvolt per metre, assuming the receiver is fed from a half wave aerial. On waves below 20 metres the noise level depends almost entirely on the noise produced inside the receiver, and hence many high-speed-telegraph services on these waves operate over long periods with an average field strength at the receiving aerial of as little as one microvolt per metre. Some provision also must be made for keeping the amplitude of the output constant for widely varying amplitudes of input, and providing selectivity. For reasons already discussed almost all commercial receivers are of the super-heterodyne type.

Whilst the general principles of receivers has remained the same, the last few years has seen a considerable advance in the technique of the circuits because of increasing knowledge in the art generally.

The selectivity of the receivers will not be governed only by the frequency band-width of the signal, but also by the frequency variations which must be allowed for, both in the transmitter and also in the heterodynes of the receiver itself. Early receivers, even those designed for the reception of C.W. telegraphy, employed intermediate-frequency amplifiers having band-widths as great as 10,000 cycles, because frequency variation was the governing factor and not signal band-width; but with the greater constancy of modern transmitters, band widths designed to suit the class of traffic being received are now incorporated in high grade receivers. Even note filters, as used in long-wave telegraph receivers, find a place in short wave work except at the highest frequencies.

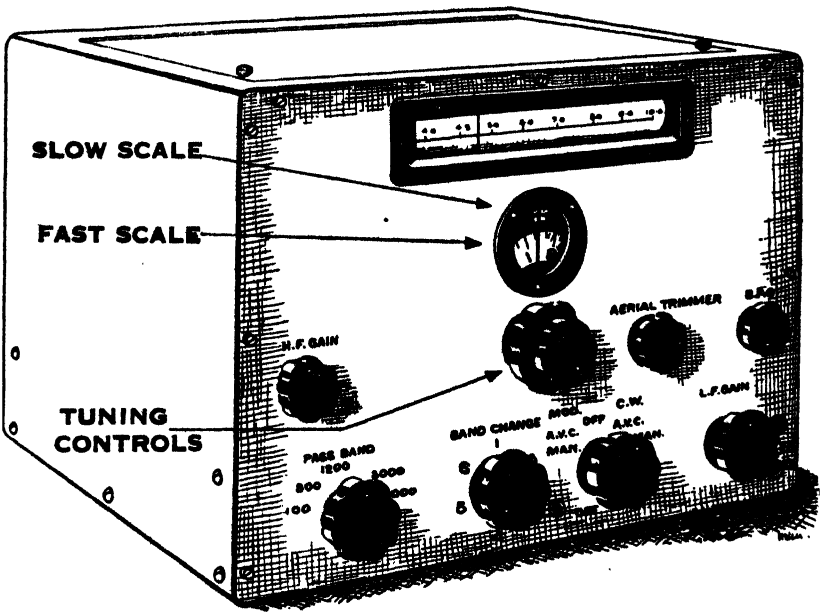
In general, the production of a constant output level on short waves can be obtained in one of two ways:

1. By arranging that the output from the weakest signal is sufficient to operate the recording gear, and then limiting the output from stronger signals.
2. By employing an automatic gain control system, in which the incoming signal controls the gain of the amplifier. In most gain controls, the rectified carrier from the last detector valve is employed to control the gain of the early high frequency stages, and is so adjusted that the weakest signal allows maximum gain.

The first system is suitable only for telegraph signals, since a limiter device would smooth out modulation. Although the receiver design must incorporate some method of levelling out the signal as indicated above, we can get further assistance by employing a diversity type aerial system; this requires special receiver arrangements which will be discussed. The principles of the super-heterodyne have already been set out in Chapter XIV, and we will describe a general purpose receiver of this type.

Description of Marconi Receiver Type CR. 100. The super-heterodyne receiver described below is typical of a

modern commercial, general-purpose receiver, and covers a wave range of 30 Mc/s (10 metres) to 60 kc/s (5,000 metres), and thus includes not only the short waveband but medium waves as well. Fig. 285 shows a general view of the receiver and it will be seen that the controls are but little more complex than those of a broadcast receiver. It is possible, however, to vary the sensitivity and selectivity between wide limits to



MARCONI RECEIVER TYPE C.R. 100

FIGURE 285.

suit the diverse conditions met with, as a receiver of this class must be capable of utilising very weak or strong signals, under very varied conditions of interference and must be able to receive C.W. telegraphy as well as telephony.

The logging scale, operated through a double slow-motion mechanism, is of interest as it provides a rapid and very accurate method of re-tuning to a station previously located. This is accomplished by combining scales geared to the hair line indicator which moves across the linear frequency scale,

these geared scales in effect developing this scale to one of 18 feet equivalent length. By such means each scale is divided into 1,250 open divisions, and as it is possible to read to one quarter of a division, even at such a high frequency as 30 Mc/s the quarter division represents less than 5 kc/s change. Thus a dial setting at the highest frequency is still sufficiently near to enable the operator to pre-set the receiver to a C.W. station to give an audible beat-note with the local oscillator. The way in which this is accomplished can be seen by reference to Fig. 285. The linear frequency scales for each range (calibrated in megacycles) are each set out lengthwise and disposed around the circumference of a barrel which is rotated by the range switch operating the various scales, the particular scale for 4 to 10 Mc/s being shown in Fig. 285 at the top of the panel.

The larger of the two tuning controls (and the faster slow motion) has attached to it an open circular scale, the lower and larger scale in the circular window, and both handle and scale rotate twenty-five times to one complete arc of condenser movement. This handle is coupled through a toothed gear (spring loaded to remove back-lash) to the main condenser spindle at a 25/1 reduction and it is also coupled by a second similar gear having a similar ratio to the outer circular scale seen at the top of the circular window. This outer scale therefore rotates at the same speed as the condenser and, like the condenser, has a total arc of travel of 180°. The disc which carries this scale is also coupled by a cable to the hair line indicator which travels across the face of the calibrated frequency scale as the condenser is moved from minimum to maximum or vice-versa. The 180° arc of the slow-moving scale is divided into 25 divisions and each half-circumference of the fast-moving scale has ten main divisions, and the provision of gearing without any back-lash makes it possible to ensure that the ten main divisions of the fast dial exactly covers one division of the slow-motion scale and thus enables the position of a station previously found to be logged exactly.

The outer and smaller handle on the tuning control knob is merely a friction driven, slow-motion handle, which can be used for fine adjustment and, since it is not coupled to any of the scales, slip and back-lash do not effect the calibration in any way.

Fig. 286 is a simplified diagram of connections and shows that the receiver comprises two radio-frequency amplifying valves (V_1, V_2), triode-hexode mixing valve (V_3), separate oscillator (V_4) I.F. amplifier (V_5, V_6, V_7) with a quartz-crystal filter, diode detector (V_8), L.F. amplifier (V_8), output stage (V_9). Automatic gain control is also provided by V_8 and a separate oscillator (V_{10}) for C.W. reception. The input to the first R.F. stage can be made either direct from an aerial through a small capacitor to the top end of the first tuned R.F. circuit, or through a coupling DD which is designed for use with a low-impedance feeder.

The R.F. or S.F. amplifier consists of two transformer-coupled amplifier stages, the complete frequency range being covered in six steps by six sets of coils. The three sets of coils associated with the two R.F. amplifier stages, together with the set of coils associated with the frequency-changer circuits, are switched and tuned simultaneously, the components of each inter-valve stage being mounted in separate screened compartments. The R.F. amplifier tuned circuits have been so designed that the gain does not change materially between the ranges. Except at the higher frequencies the impedance of the first circuit is such as to ensure that the thermal-agitation noise of this circuit is above the valve shot-noise. Under these conditions the signal/noise ratio is the maximum that can be obtained.

The R.F. circuit output is fed into the control grid of a triode-hexode frequency-changer V_3, but the triode section of the valve is not used as an oscillator (for reasons already explained in Chapter XIV) but the output of a separate oscillator V_4 is coupled into the grid of the triode part of V_3.

The I.F. amplifier has three valve stages working at a frequency of 465 kc/s. There are eight tuned circuits and a crystal filter (Q in the diagram) and a switch controls the width of the passband, between limits determined by the class of intelligence being received. The switch provides five different widths of passband, namely, 6,000, 3,000, 1,200, 300 and 100 c/s., but it should be explained that only the first four of these are controlled solely by the I.F. characteristics. The 100 c/s band width is achieved by audio-frequency circuits described later. For the 6,000 c/s and 3,000 c/s passbands the crystal

filter is inoperative, the change of band-width being obtained by varying the couplings. The 1,200 c/s and 300 c/s band widths are obtained by the use of the crystal filter circuit, alteration of the band-width being achieved by changing the phase of the side-circuit tuning.

The I.F. amplifier is designed to give a symmetrical selectivity curve for all the above passbands and the gain does not change by more than 6 db. whatever the position of the passband switch.

It will be appreciated that the 100 c/s band-width can only be employed for the reception of C.W. signals, when heterodyned by the beat oscillator in order to give a 1,000 c/s output to which the audio-frequency circuits are then tuned. The 100 c/s band-width will only be suitable for reception from a very steady transmitter but will enable such signals to be read through very severe interference.

Diode detection is employed, using the left-hand diode of V_8 Fig. 286. The network of resistances and condensers between this diode and the cathode form the diode load and also a filter to eliminate the I.F. The grid of the triode portion of V_8 is supplied with the detector output through a potentiometer which provides the L.F. gain adjustment.

The triode portion of V_8 is resistance-capacity coupled to V_9, the filter being only in circuit when the passband switch is set at 100 c/s. The circuit is so arranged that the L.F. amplification is nearly the same, whether the filter is in use or not.

The output valve (V_9) is of the "beam-tetrode" type and its output can be delivered in three ways:

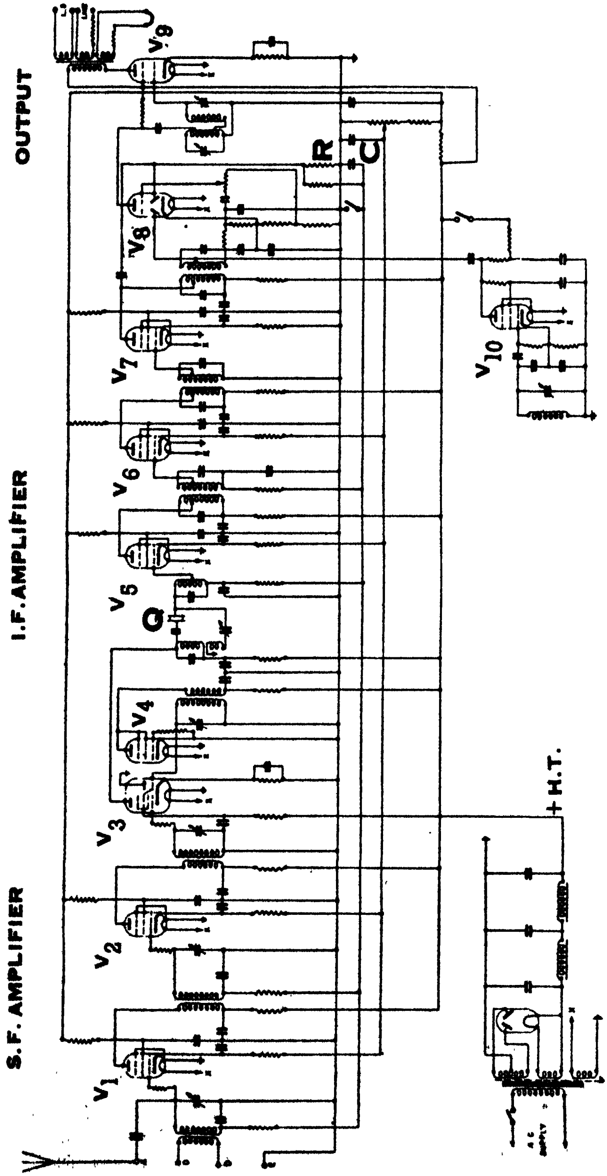
(*a*) To headphones, when the output is limited to about 15 db below 1mW (a sufficiently low level not to overload the ear).

(*b*) To a low-impedance loud-speaker, an input of 2W to a 2 Ω speaker being possible.

(*c*) To a 600 Ω line, the maximum output being 2mW.

Output (*a*) can be used at the same time as (*b*) or (*c*) in order to monitor and it is also possible to use (*b*) and (*c*) as additional headphone outputs if desired.

A manual control of H.F. gain is provided which varies



MARCONI RECEIVER TYPE C.R. 100

Figure 286.

the cathode potential (and therefore the grid bias) of V_1, V_2, V_5, and V_6.

The control voltage for A.G.C. is derived from the primary of the last I.F. transformer and is rectified by the right-hand diode of V_8. The voltage developed across R is applied to V_1, V_2, V_5 and V_6. Since the receiver has to work on both telephony and telegraphy, it is necessary to vary the time-constant of the A.G.C. circuit by varying R and C. For telegraphic reception the time-constant must be long enough to prevent noise surging up to a high level during the intervals between Morse "marks," even at hand-signalling speeds.

The various changes necessary when changing from telephony to telegraphy are all made by one knob which also cuts out the A.G.C. when required.

The gain of the receiver and the design of the A.G.C. system ensures that inputs capable of providing a signal/noise ratio of 10 db. operate the A.G.C. system. For inputs greater than the above, the A.G.C. decreases the gain of the receiver so that its output does not increase appreciably, but the decreased gain of the receiver ensures a decreased output of noise, with a consequent improvement in signal/noise ratio.

Discussion of CR 100 Performance. We propose now to discuss very briefly how far a receiver of this type may be expected to approach the performance ideally desired.

As has been explained in a previous chapter, inherent noise of a receiver, either shot-noise or thermal-agitation noise, sets a limit to the useful gain of a receiver. An unqualified statement of the overall gain of a receiver from radio-frequency input to voice-frequency output is therefore an insufficient definition of performance. It is accordingly usual to specify a receiver performance in terms of the radio-frequency inputs necessary to give a stated signal/noise ratio.

Table I shows, for the CR 100 receiver, the inputs necessary (at four selected frequencies) to give the stated signal/noise ratios.

TABLE I

Input Radio Frequency mc/s.	Metres Wavelength	μV Input (3 kc/s passband)	
		Signal Modulated 30% (10 db. S/N Ratio	Unmodulated Signal (20 db. S/N Ratio)
1·4	214	2·0	2·0
4·0	75	2·0	2·0
11·0	27	2·0	2·0
28·0	10·7	3·0	3·0

The figures quoted are taken with a signal generator through a load of 100 ohms pure resistance, this latter figure representing the input impedance of the receiver. For modulated inputs the input to the receiver is increased until the difference between the outputs with modulation on and off is 10 db. For unmodulated inputs the beat-frequency oscillator in the receiver is employed. With the signal generator connected up in order to introduce the damping of the fictitious aerial, but with the signal switched off or mistuned sufficiently to contribute nothing to the receiver output, the voice-frequency gain control of the receiver is adjusted to give a noise output at least 20 db. below the maximum receiver output. The signal is then correctly tuned, and its input value adjusted to give an output 20 db. above that of noise.

In considering the relation of the above figures to the ideal desired, i.e. maximum possible signal/noise ratios, two facts must be borne in mind.

(*a*) The shot-noise of a frequency-changer valve is inherently worse than that of a straight amplifier valve, and it is therefore desirable to employ sufficient radio-frequency amplification to ensure that noise occurring before the frequency-changer valves is the limiting factor. Except for a small portion of the highest frequency-range this is achieved in the receiver under discussion.

(*b*) Secondly, the thermal circuit-noise of the first tuned circuit in the receiver should be greater than the shot-

noise of the first amplifier valve. This is desirable because increase in dynamic impedance of the first circuit permits increase in the transfer ratio from the receiver input to the first grid. A limit is however reached when the dynamic impedance of the first circuit is so high that its thermal

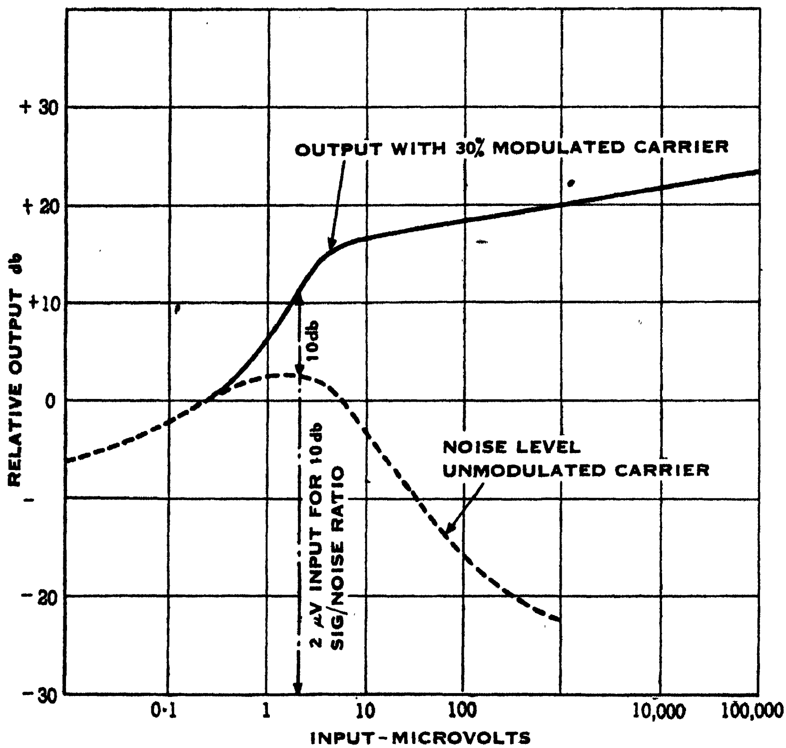


FIGURE 287.

noise controls the useful gain of the receiver. Any further increase in dynamic impedance would increase the transfer ratio from receiver input to first grid but this increase would be accompanied by a pro-rata increase of noise, and no advantage would result.

From this point of view the receiver under discussion does not meet ideal requirements at the higher frequencies, since it is impossible to obtain the necessary dynamic impedance

for the first tuned circuit. Below 3 Mc/s, however, the required ideal is satisfied.

Considering the A.G.C. system, the performance of this must be considered in relation to useful sensitivity. Otherwise, a receiver may give a quite satisfactory signal/noise ratio with an input of say 5 μV, while the A.G.C. may not commence to operate until inputs of 50 μV or more are applied. This state of affairs permits wide variations of output level to occur at usable signal inputs.

The general considerations above outlined in respect of signal/noise ratio and A.G.C. operation make it useful to depict receiver sensitivity by a curve such as Fig. 287 taken at the frequency giving the worst performance. From this it is possible to read off the input required for any signal/noise ratio. It also shows the extent to which the A.G.C. system is linear, and on this depends the progressive improvement in signal/noise ratio with increased input.

In this receiver, the radio-frequency tuning attenuates the image frequency by 30 db. on the highest frequency band and by 100 db. on the lowest.

It was explained on page 434 that the selectivity of any receiver can be specified by three parameters and Table II shows the values of these for the CR 100 receiver. Since the adjacent channel selectivity is due to the I.F. circuits, these values will apply whatever the frequency being received.

TABLE II

Band-Pass Frequency. kc/s.	–6 db.	Frequency spread (kc/s) (kc/s) at –40 db.	–60 db.
6	6	16	21
3	3	8·5	13
1·2	1·2	5·4	8·8
·30	·3	4·5	8·8

Diversity Reception. It has been found, experimentally, when recording high-speed telegraph signals, or telephony, that the sudden "deep fades" that are so troublesome are very local; that is, when the signal from a station is at a very low level in one spot at a certain instant, it may be quite high in another spot only a few hundred feet away.

In consequence of this fact, considerable work has been done on combining the signal received from several aerials so as to produce a more uniform output level and eliminate as many deep fades as possible.

The problem of "mixing" the E.M.F.s given by the various aerials is quite different from that encountered when spaced aerials are used to give directional properties, for in this case the relative phases of each aerial must be retained, as these determine the directive properties. In diversity reception we wish to combine in a scalar fashion the outputs from each aerial (or array), irrespective of the phase differences which must exist due to spacing. There are at least three possible schemes.

(1) Use a separate receiver for each array and combine their low-frequency outputs. The disadvantage of this method is that when a deep fade occurs on any array the receiver connected to it gives no signal output, but still gives its full noise output, and therefore the resultant signal/noise ratio suffers. Another disadvantage is that if any of the arrays are receiving a very distorted signal, due to the selective fading of the carrier frequency at any instant, then this array will contribute distortion to the final result.

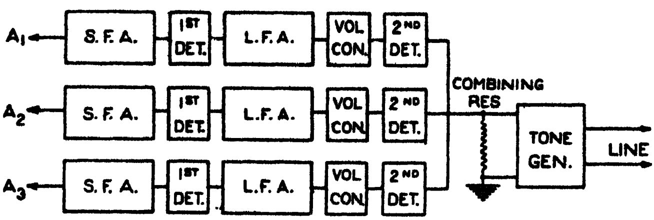
(2) Use a separate receiver for each array as before, but automatically select the signal mainly from the receiver giving the best output. This evidently avoids the faults of (1).

(3) After some separate amplification, use a rotating switch (or its equivalent in a valve network), which will apply the arrays in rapid succession to the remainder of the receiver. In order to be effective, this switching must connect each array at least once during each Morse dot, and hence is really modulating the signal. This necessitates a wider frequency band in the receiver, and this in turn increases the noise level.

It is generally agreed that the use of some diversity method of reception is desirable, but it does not, as is often supposed, provide a simple and cheap system.

Not only must two or more receivers per circuit be provided, but if a good service is required, the diversity aerial cannot be of the simple omni-directive type, but each must comprise in itself a small array.

Radio Corporation of America's Diversity Receivers. The R.C.A. has used the diversity principle widely, and in Fig. 288 is shown schematically the arrangement of their telegraph receiver. The arrays usually used have been described on page 223. Each receiver embodies tuned amplification at the signal frequency, using screen-grid valves. This is followed by a self-oscillating detector to produce heterodyne beats. The battery supplies for these detectors are separate from those of the remainder of the receiver, in order to assist



SCHEME OF R.C.A. TELEGRAPH RECEIVER.

FIGURE 288.

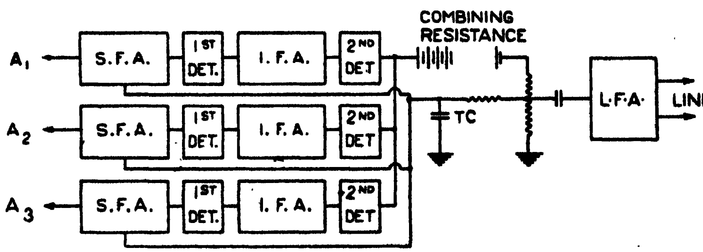
in maintaining the frequency constant. A two-stage selective low-frequency amplifier follows, having a band-width from 100 to 7,000 cycles, and means are then provided for manual volume control, and the gain of each receiver is here adjusted to be about the same value. The low frequency is now rectified to D.C., and the outputs of each receiver combined to produce a resultant drop in the combining resistance shown. This voltage then operates a tone generator. It will be seen that this receiver works on method (1) of the previous section.

A somewhat different type of equipment has been developed for telephony signals, and has been applied particularly to the re-broadcasting of European broadcast transmissions through American stations.

In this case three separate super-heterodyne receivers are provided, shown schematically in Fig. 289. The detectors at the end of the I.F.A. are supplied from a common battery through

a common resistance. The A.C. component of the voltage produced by the rectified signals is passed on to a low-frequency amplifier, and thence to line. The D.C. component of the voltage is applied (through a time-constant circuit) to the grids of the high-frequency amplifier, thus controlling the gain.

It will be seen that the arrival of a strong signal on one of the arrays decreases the gain of all the S.F.A.s (Signal Frequency Amplifiers) and therefore reduces the noise contributed by the other two. The second detectors are high-amplification valves used with considerable negative bias so that the output is approximately proportional to the square of the input. Thus the receiver having (say) twice the strength of the others (and



SCHEME OF R.C.A. TELEPHONE RECEIVER.

FIGURE 289.

therefore having probably the best quality speech), will contribute four times the output of the others.

Diversity Reception at the Somerton Station of Cable and Wireless Ltd. Diversity reception has also been brought into considerable use at the above station, and a long series of tests made to determine its effectiveness and the best arrangement to use.

The usual method is to employ two arrays (see pages 215 and 233) spaced broadside on to the incoming signal. Ordinary types of receivers are connected to each array, and their output combined by the circuit of Fig. 290. Before using diversity reception, it had been found that the use of a grid resistance (R_1) in the detector circuit was useful to assist in maintaining constant output. Any increase of signal above the value that makes the grid voltage of the detector zero produces grid current, and R_1 then causes a negative bias, so reducing the effect of the stronger

signal, and the voltage developed across the resistance in the common anode lead is therefore nearly constant for all signals above a certain strength. This is, of course, a form of automatic gain control, and the time constant of C_1R_1 must maintain the bias during the spaces in the Morse.

The additional bias applied to both detectors therefore depends upon the sum of the signals received on both aerials, so that if one receives a strong signal, the other is so much "backed

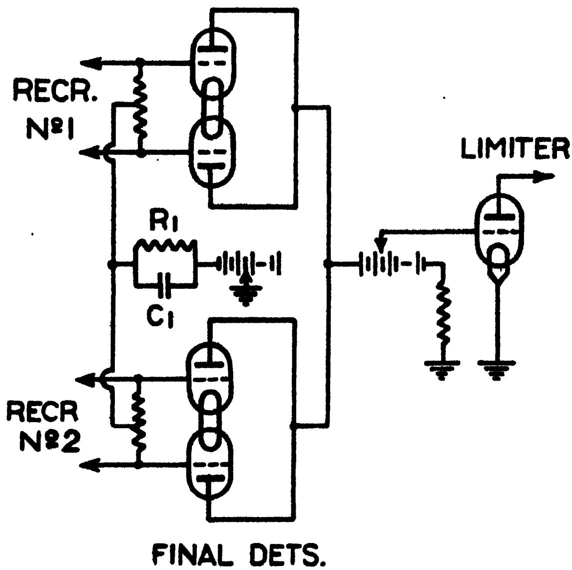


FIGURE 290.

off" as to contribute nothing to the result. This combining circuit is therefore of the second type discussed on p. 464.

The valve-circuit labelled "Limiter" in Fig. 290 is often used in telegraph receivers. It will be seen that a detector current of a certain value will provide sufficient voltage drop across the resistance to reduce the anode current to zero. Clearly, a larger detector output can do no more and therefore the signal passed on to the recording apparatus by the limiter cannot exceed a limiting value in spite of large fluctuations of received signal-strength.

It has been found that diversity reception is most useful

when pure C.W. is emitted by the transmitter rather than when I.C.W. is used. This might be expected because if I.C.W. is used, the transmission occupies a band of frequencies and this in itself tends to overcome the type of fading which diversity is designed to combat.

An extended range of tests have been carried out to determine the best spacing to use, and also the minimum spacing which will give useful results. It appears that the spacing required (expressed in wavelengths) falls as the wavelength is increased —a very fortunate result—since it means more or less equal actual spacings for the different wavelengths.

The results obtained on the New York circuits may be summarised thus:

Wavelength (Metres).	Best Spacing.	Minimum Useful Spacing.
[1] 15·89	20λ	7λ
22	10λ (6λ good)	4λ
40·5	7λ to 10λ (varies)	4λ (good)

Spacings are between centres of each array.

As a result of the observations, it has been concluded that if space is allowed in laying out a station for a 30λ spacing on 14·5 metres, there will be ample room for suitable spacings on all other wavelengths comprised within the short wave band.

The diversity principle has been found of the utmost value when signals are strong (in average value), but suffering from severe rapid fading. Under these conditions even two single aerials suitably spaced may be nearly as good as a single large array. On the other hand, when signals are weak, the much larger "pick-up" of the array is most valuable, enabling traffic to be carried on, whilst a single aerial gives a barely audible signal at the receiver output. It would appear, therefore, that the soundest and most economical receiving system should consist of medium-sized arrays suitably spaced. The use of a very large receiving array does not appear desirable, since its polar diagram is very sharp, whilst the direction of incoming signals is liable to change under some conditions, and also the phase of the incoming signal may not be the same over the whole width of the array.

[1] In this case each array was a Franklin Beam (Uniform aerial type), 3·4λ wide.

CHAPTER XVI

COMMERCIAL WIRELESS TELEPHONE CIRCUITS

IN recent years the telephone networks of the principal countries have been linked up by international circuits. Where long distances across the ocean have to be bridged, wireless is exclusively used, and it is to these channels that we have applied the term "Commercial Wireless Telephone Circuits."

General Requirements of a Commercial Wireless Telephone Circuit. The ideal to be aimed at is, of course, that subscribers using such a circuit should not be reminded of its character or of its length by any special peculiarities or by poor performance. The output should, therefore, be at a constant and sufficiently high level, there should be an absence of noise and echo and the circuit should pass satisfactorily frequencies between 250 and 2750 cycles per sec. Further, the wireless circuit must be so terminated that the usual subscribers' instruments may be used in the normal manner.

The wireless circuit has two undesirable features which will make the attainment of this ideal difficult. First, a signal/noise ratio poorer than that of a trunk line (because the use of intermediate repeaters is impossible) and secondly, a transmission path which is very variable compared with a well-maintained line circuit, so that varying attenuation and possibly distortion is introduced.

In order to obtain a satisfactory signal/noise ratio, highly directional receiving systems will be essential and, on short waves, directional transmission will also be used. Whilst a satisfactory signal/noise ratio can be obtained on a 5000 metre wavelength across the Atlantic by the use of a single sideband suppressed carrier transmitter with an input of about 200 kW and a directional receiving system, it is not economically possible to use long waves over greater distances or in direc-

tions other than away from the prevailing atmospheric centre, and short waves are invariably used.

The varying attenuation will have to be corrected for by an efficient automatic gain control in the receiver. Special methods of working are being introduced to reduce distortion.

Terminal equipment of a special character will be necessary in order that the wireless circuit may be connected to an ordinary telephone circuit and considerable amplification will need to be available at the terminal points so that losses in the land-line extensions may be made good and the wireless transmitter kept fully modulated, which is essential in order to maintain the signal/noise ratio.

The "transmission time," that is, the time taken for a given portion of the speech wave to travel from one subscriber to the other, should be as small as possible. This "transmission time" must not exceed 0·5 sec., otherwise more than one second will elapse when a speaker has finished talking before he can receive any reply. This would cause difficulty to those not used to the circuit and both subscribers would probably speak at once. The wireless circuit has here a great advantage since the speed of the actual wireless wave is 300,000 kms. per sec., as against about 70,000 kms. per sec., for an average trunk line. It is necessary, however, to see that the overall transmission time is not increased by the circuits incorporated in the terminal equipment especially in view of the fact that the wireless links are often connected to long trunk lines in order to reach another country.

The wireless circuit should be "secret," that is, it should be as difficult for an unauthorised person to "eavesdrop" on a conversation involving a wireless circuit as it would on one over a line circuit. Owing to the common use of the intervening medium this secrecy can only be obtained by the use of some privacy system which will render the speech unintelligible to all except those possessed of elaborate equipment plus special information.

The principles of suitable transmitters, receivers and array systems are described in the appropriate chapters of this book, and this chapter will deal only with terminal apparatus, privacy apparatus and some special methods of working.

The Four-Wire Line Circuit. Before considering the

wireless circuit further it may be helpful to consider briefly the working of a long distance line circuit since the wireless circuit is really a special case of this.

A true duplex system is one in which the speech currents are free to travel in both directions at the same time. If these currents travel over the same line and are merely separated at each end by the combined transmitting-receiving apparatus, we have what is known as a two-wire circuit.

It will be understood from what is said in Chapter VI, that if the impedance of the apparatus at the ends of the line is not the same as the characteristic impedance of the line, there will be reflection and when this reflected wave reaches the speaker's receiver, he will hear an echo of his speech.

If the line is short, the time before the echo returns will be so short that the subscriber will not distinguish it from the "side-tone" of his own speech which he always hears in his receiver and he will not, therefore, be troubled by the echo.

When the channel has to be extended to greater distances it will be necessary to include repeaters, that is, valve amplifiers, in the circuit. There are two ways in which this may be done:

(1) By the insertion of two-way repeaters at suitable intervals into the two-wire line.

(2) By connecting the local, short, two-wire line to a four-wire line so that the "go" currents travel by one pair of conductors, and the "return" currents travel back by a second pair of conductors. One-way repeaters can then be inserted in the four-wire portions of the circuit.

Since the valve amplifier is essentially a "one-way" device, the two-way repeater used in the first method involves two amplifiers and balancing networks to separate out the speech currents travelling in the two directions. Unless these balancing networks are very carefully adjusted, an echo will be returned from each one and also "singing" (that is, self-oscillation) may occur. The arrangement, therefore, becomes progressively more difficult to set up and adjust as the number of repeaters and their amplification is increased, and the two-wire arrangement is rarely used in practice where more than two or three repeaters are to be inserted in the line.

The second method is shown diagrammatically in Fig. 291, where S_1 and S_2 are the subscribers, TE_1 TE_2 balancing networks at the terminal points and R,R, represent simple "one-way" repeaters. The centre portion of the diagram, within the chain dotted lines, will be considered later and should be ignored at this stage.

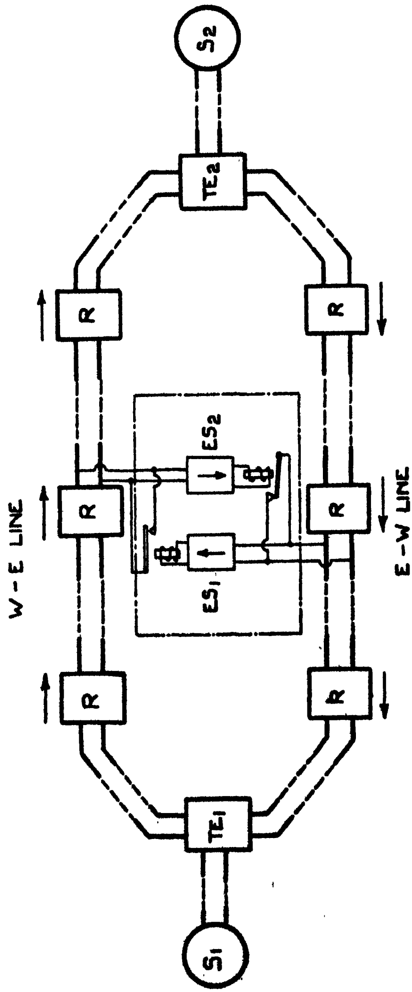


FIGURE 291.

Four-Wire Line Terminal Equipment. The first requirement at TE_1 and TE_2 will be a differential transformer termed by telephone engineers a "hybrid coil," the principle of one type of which is illustrated by Fig. 292, shown connecting a "go" and "return" line with a local subscriber. In this diagram all the coils are assumed to be wound in the same direction. Speech currents are presumed to be incoming from the distant subscriber and an equivalent alternator is shown whilst the local subscribers apparatus is represented by the Z at the end of the two-wire line. Suppose the direction of the E.M.F.'s induced in MM' and OO' by the current in PP' and QQ' to be in the direction (at the instant chosen) of the full arrows. Then currents flow in the two-wire line and in the balancing network in the directions shown by the dotted arrows. If the impedance of these circuits is the same, then these currents are equal and induce equal

E.M.F.'s in JJ' and KK' but since the currents flow through LL' and NN' in opposite directions, the induced E.M.F.'s cancel out and no current flows in the "go" line. Half of the energy from the "return" line is, of course, lost, since equal currents flow in the two-wire line and its balancing network.

If now the alternator is placed at the local subscriber's position, then a current will flow in LL' and $M'M$ inducing E.M.F.'s in both JJ', KK' and PP' QQ', but there is not a

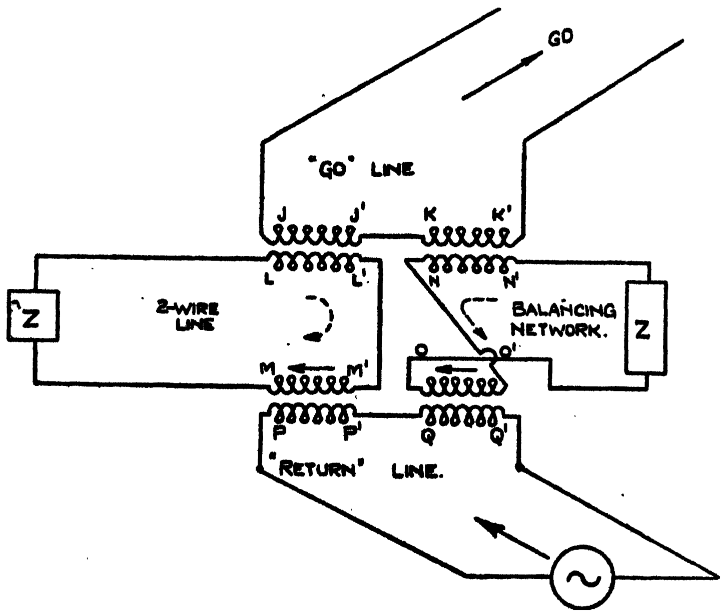


FIGURE 292.

complete path through the return line because there will be valve amplifiers in this line "facing the other way." It should be noted, however, that although no current flows in the balancing network, half the energy is again lost, this time in the "return" line valve anode impedance.

An alternative form of hybrid coil is shown in Fig. 293a. Again considering an input from the return line, an E.M.F. is induced in LL' and KK', and this circulates a current through the balancing network, and the two-wire line, in series. Since the combined impedance of M_2L', the balancing network, and $K'M_1$ is equal to the combined impedance of

M_1K, the two-wire line, and LM_2 and equal E.M.F.'s are induced in each half winding, there will be no P.D. between M_1 and M_2. If necessary, reference to a similar circuit (Fig. 293b) will make the matter clear.

The first type of hybrid coil is somewhat more costly in material than the second but each winding can be manufactured to wider tolerance and yet a good balance obtained by suitable pairing up of the coils after test.

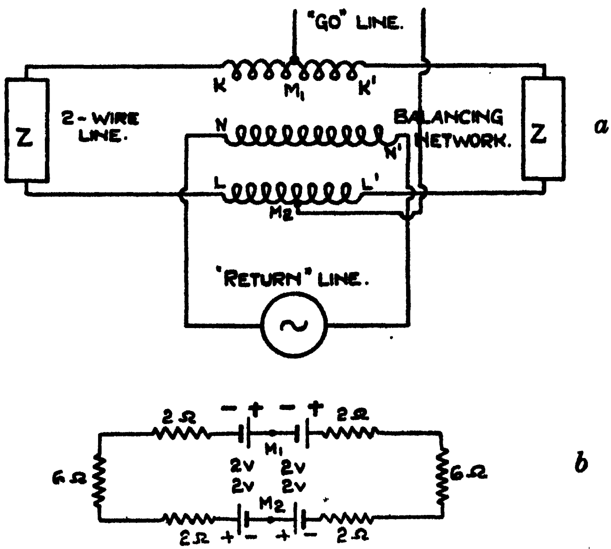


FIGURE 293.

It might be thought that this provides all that is required for satisfactory operation but difficulties would arise in practice owing to imperfect balance. Suppose that the windings of the hybrid coil are not perfectly balanced or that the balancing network does not exactly simulate the impedance of the two-wire circuit, then a speech wave arriving from S_2, Fig. 291, will be partly transferred at TE_1 to the W—E line and returned to (TE_2) and the greater portion of this "straying" speech wave will reach S_2, so that S_2 hears his own speech as an echo and if the transmission time be long this will be troublesome. If there is also unbalance at (TE_2), some of the wave will return again via the E—W line to (TE_1) and the greater part

thence to S_1 who will, therefore, hear an echo of S_2's speech approximately twice the transmission time after hearing the original speech.

Further, even if the terminal hybrids are perfectly balanced, should the two-wire circuit to S_1, for example, not be perfectly terminated, reflection of speech received from S_2 will occur and when this reflected wave reaches the hybrid coil at (TE_1) it will be passed over to TE_2 and eventually reach S_2 as an echo of his own speech, since even a perfect hybrid coil cannot distinguish between S_2's speech reflected from S_1 and S_1's own speech.

In addition to the inconvenience of these echoes it is evident that should the gain around the circuit (TE_1), W—E line, (TE_2), E—W line, (TE_1) exceed the loss, an oscillation would commence, making speech over the circuit impossible. Such an oscillation is usually referred to by telephone engineers as "singing" and the adjustment at which it commences, the "singing" point.

It will be seen, therefore, that the use of a hybrid coil alone would require a perfectly constructed coil, perfectly balanced networks, and perfectly terminated lines, so that whenever length or characteristics of the two-wire circuit were changed at all a careful re-adjustment of the balancing network would be necessary.

These difficulties are overcome in long four-wire line circuits by working with gains which keep the circuit below the "singing" point, and by the use of "echo suppressors" by means of which the speech currents from one subscriber render transmission in the reverse direction impossible while they last. The channel is, therefore, no longer truly duplex but the subscribers are not usually conscious of anything peculiar in the operation of the channel.

One arrangement of an Echo Suppressor near the middle of a long four-wire line circuit is shown schematically in Fig. 291. ES_1 and ES_2 are amplifiers which energise relays which, on closing, short circuit the lines. When neither subscriber is speaking both lines are complete, but should subscriber S_1 speak, for example, a portion of his speech wave, on reaching the repeater station half-way along the line, is amplified by ES_2 and caused to close the contact short circuiting the E—W

line. It will be seen that the relay need not be particularly rapid in its operation, all that is necessary is that it shall close before the echo reaches half-way back along the line.

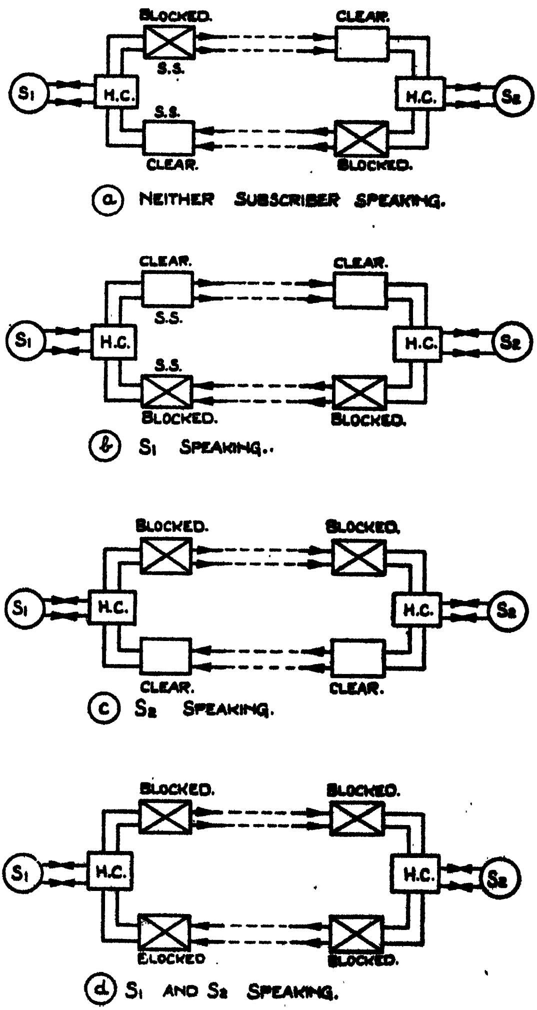
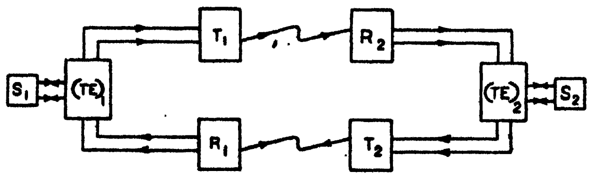


FIGURE 294.

It will be evident that with any circuit having echo suppressors only it is not possible for the gains of the repeaters to be adjusted to equal or exceed the losses in the lines plus the loss across the hybrids from "return" to "go" lines, otherwise a

complete circuit having zero or negative resistance and, therefore, capable of self-oscillation will be formed when neither subscriber is speaking. The "singing" would be intermittent because its onset would cause the relays to close but it is obvious that the circuit would not be workable.

Should we desire to have an overall gain round our four-wire circuit in order that the output into the distant two-wire circuit may be considerably greater than the input (in order, perhaps, to obtain a sufficient strength at the distant end of a poor, two-wire line) we shall need to convert our Echo Suppressor into a Singing Suppressor which always keeps one



SCHEME OF TELEPHONE CIRCUIT

FIGURE 295

line blocked even if neither subscriber is speaking. The action of a Singing Suppressor is shown schematically in Fig. 294.

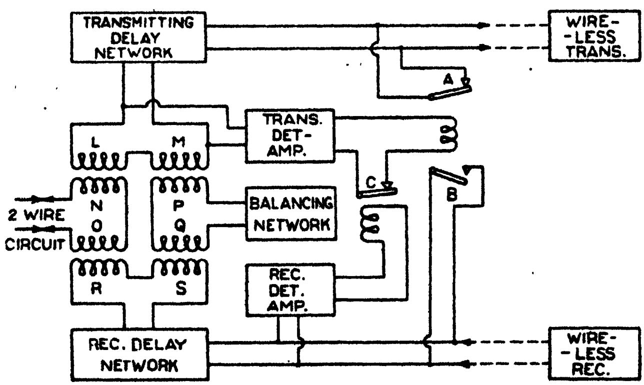
Such a circuit employs two interlocked suppressors at each end, one in the "go" line and one in the "return" line. In the quiescent condition both go lines are blocked, and both return lines are clear.

Speech from a distant subscriber, S_1 say, to one terminal TE_1 operates the interlocked suppressors there, blocking the return line and clearing the go line. Since the other end is already clear, the go currents will reach the subscriber at S_2. The action is reversed when S_2 speaks, but in cases when both subscribers speak together, all four suppressors become blocked. The circuit changes must, of course, be made automatically without the subscribers being aware of anything unusual in the behaviour of the circuit.

The wireless telephone is a special case of such a four-wire line circuit just described, as it invariably works at a large overall gain, but it differs therefrom in that the attenuation

is much more variable and the signal/noise ratio usually poorer, and the singing suppressors used will need to be of a more elaborate type to cope with the greater variety of service required of them.

Fig. 295 shows schematically a wireless telephone circuit where TE_1 TE_2 are the terminal equipments, and T_1 and R_2 the transmitters and receivers at each end. It will be observed that in addition to possible "sing" around the main loop circuit, there are also loop circuits at each end from T_1 to R_1 and T_2 to R_2, and should the circuit be operated at the same



TELEPHONE TERMINAL EQUIPMENT.

FIGURE 296.

wireless frequency in both directions "sing" around the local circuits must also be prevented.

Three types of terminal equipment designed to work on wireless circuits will now be described.

Singing Suppressor—Relay Type.[1] The arrangement first developed by the A. T. & T. Corp., for four-wire line circuits was modified for use on the trans-atlantic long wave circuit. The simplified scheme of connection is shown in Fig. 296.

Suppose speech to be incoming from the two-wire line. Currents flow in L and M, and are amplified and rectified by the transmitter amplifier-detector and first close the relay B, thus shorting the receiver line. A is then opened, thus making the transmitter available. When speech is incoming from the wireless receiver B will be opened and A closed. The received

rectified signal will cause C to open and thereby ensuring that even if the "near" subscriber speaks at the same time, A cannot open or B close.

It will be appreciated that since the first syllable of speech has to work the relays, they must either be very sensitive and quick acting—in which case they are liable to false operation due to noise—or some speech must be lost. To overcome this difficulty the delay networks are inserted. These are so proportioned that the voice currents take a comparatively long time to travel through them, and hence the relay has time to function and need not be unduly sensitive.

It has been already explained that the transmission time should, if possible, not be increased because this limits the length of line which can be connected to the ends of the wireless circuit. To overcome this and also to obviate the adjustment of the relays an entirely valve operated arrangement has been developed by the British Post Office and is in use in all the wireless terminal equipment at the London International Exchange. It has the additional advantage that a differential arrangement to prevent false operation due to noise from the receiver is incorporated.

British Post Office Terminal Equipment. A simplified circuit diagram is given in Fig. 297. This diagram includes the privacy equipment which requires additional suppressors and an extra hybrid coil, but in the preliminary explanation this will be ignored, and we can assume that the input from the wireless receiver is connected to the points AB, and the speech output to the modulator of the transmitter from the points CD direct.

Examination of this diagram shows that it includes a suppressor in the receiving line RS_1, a suppressor in the transmitting line TS_1, a main suppressor which controls RS_1, and an auxiliary suppressor which normally blocks TS_1 unless it itself is blocked. In considering the working of the system we must keep in mind the general conditions set out previously, namely, that with no speech passing either way the incoming line from the receiver is open, and the outgoing line is blocked. Further, incoming speech maintains these conditions, but outgoing speech must be capable of reversing the conditions. The suppressors are the means of accomplishing this, and each

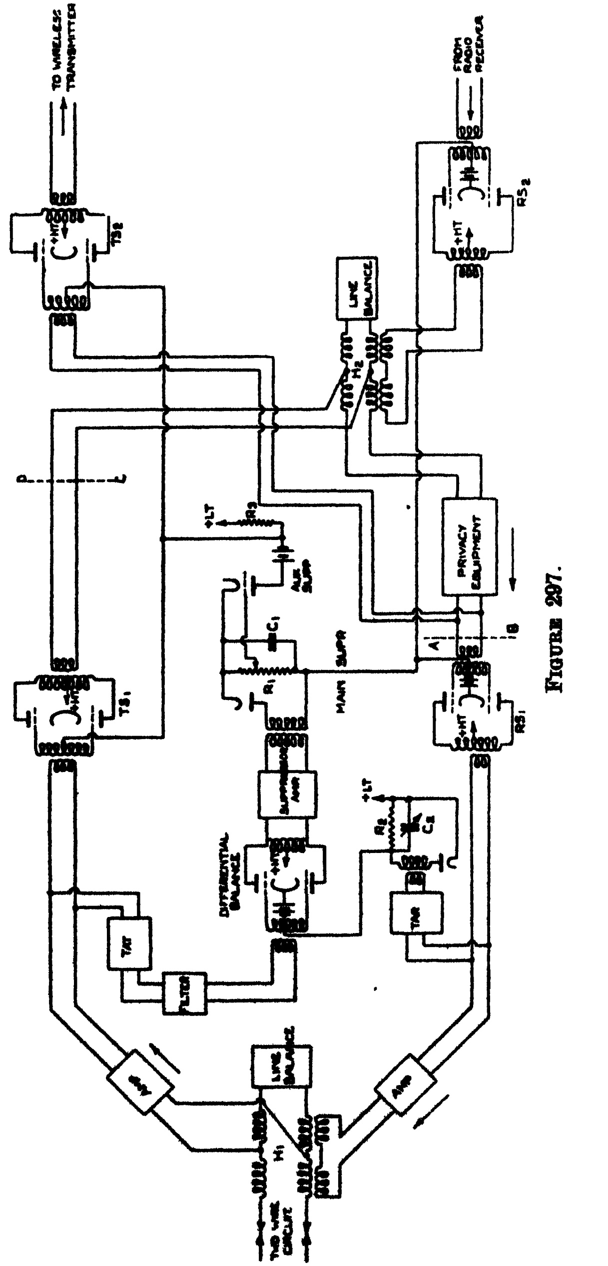

TO WIRELESS TRANSMITTER
FROM RADIO RECEIVER
LINE BALANCE
PRIVACY EQUIPMENT
MAIN SUPR
AUX SUPR
SUPPRESSOR AMP
DIFFERENTIAL BALANCE
TAT
TAR
FILTER
AMP
TWO WIRE CIRCUIT

FIGURE 297.

consists of a single push-pull valve stage having zero gain in the clear condition and about 100 dbs. loss when blocked by negative applied to their grids.

It will be observed that the anode circuit of the auxiliary suppressor is coupled to the grid of the transmitting suppressor TS_1. This means that when the auxiliary suppressor passes current, namely when its grid is zero, the anode current in this auxiliary suppressor applies negative to the grids of the transmitting suppressors TS_1. Now the auxiliary suppressor is controlled from the resistance R_1, the potential across which in turn is controlled from the differential balance, connected to the " go " and " return " lines through amplifiers TAR and TAT. Thus when no speech is passing in either direction (and assuming that no noise is present either), because no E.M.F.'s are applied to the amplifiers TAR and TAT, the suppressor amplifier applies no E.M.F. to the main suppressor and no current passes through the resistance R_1. Thus no negative E.M.F. is applied to the grid of the auxiliary suppressor (hence it has anode current), or to the grid of the main receiving suppressor RS_1, and in consequence this remains clear, but because of current through the auxiliary suppressor valve it applies negative to the transmitting suppressor TS_1 and effectively blocks the outgoing line. Thus in the quiescent condition the receiving path is open and that to the transmitter blocked. If the distant subscriber now speaks, the conditions remain the same except that speech voltage is applied to the amplifier TAR, which backs off the grids of the differential balance, but this makes no difference to the operation of the circuit. At the same time, however, if the near subscriber also speaks, because the incoming speech has backed off the differential balance more negative, the transmitter speech is unable to pass current through the differential balance if the adjustments have been made correctly and in consequence produce current across the resistance R_1.

On the incoming speech ceasing, the condenser C_2 discharges comparatively slowly, removing the bias from the differential balance. This is necessary in order that echoes returning from the end of the two-wire circuit may not change the suppressors over to the transmitting condition. C_2 may be varied for long, medium, or short two-wire lines.

Suppose now that only the near subscriber speaks; C_2 having discharged, there will be no bias produced by *TAR* and hence a portion of the outgoing speech current passes through the differential balance and a steady P.D. is produced across the resistance R_1. This P.D. applies sufficient negative bias to RS_1 to render it inoperative. Bias is also applied to the auxiliary suppressor and hence there is no longer any P.D. across R_3 and, therefore, no bias on TS_1, and this suppressor is, therefore, clear for outgoing speech to the wireless transmitter.

We have now to consider how the performance of the circuit is modified by the presence of noise (especially that incoming from the wireless receiver) and by unbalance at the hybrid. Suppose neither subscriber to be speaking, then noise will be coming in on the receiving line and some of this will pass across the hybrid coil due to unbalance. Although the loss across the hybrid will be high, yet, as there are two amplifiers in series (each of which may be adjusted, under certain circumstances, to give a high gain) sufficient energy might pass through *TAT* to produce a large enough P.D. across R_1 to cause the circuits to change over to the transmitting condition.

This is prevented, however, by the fact that some of the received noise is passed through *TAR* and produces a P.D. across R_2, thus biassing the differential balance. Hence by suitable adjustments of *TAR* and *TAT*, the biassing of the differential balance can be made sufficient to prevent operation by noise through *TAT* and yet to permit of operation by outgoing speech. If now the distant subscriber speaks the bias produced by R_2 is increased and as some of the speech energy will also leak across the hybrid, the energy passing through *TAT* will also be increased. The adjustments must be such that receiver noise plus received speech produce sufficient bias across R_2 to lock the differential balance and keep the suppressors in the receive condition. As the incoming noise increases it will be necessary to reduce the gain of *TAR* in order that outgoing speech may not fail to change over the suppressors. Finally, a limit is reached where false operation by noise cannot be prevented, but before this the circuit would probably be "uncommercial" due to very poor signal/noise ratio.

In order to assist further in preventing operation by noise

(especially that coming from the two-wire circuit) *TAT* is followed by a filter passing only frequencies from about 500–2,500 cycles so that low frequency noise is cut out.

This shunting condenser C_1 gives the combination R_1C_1 a

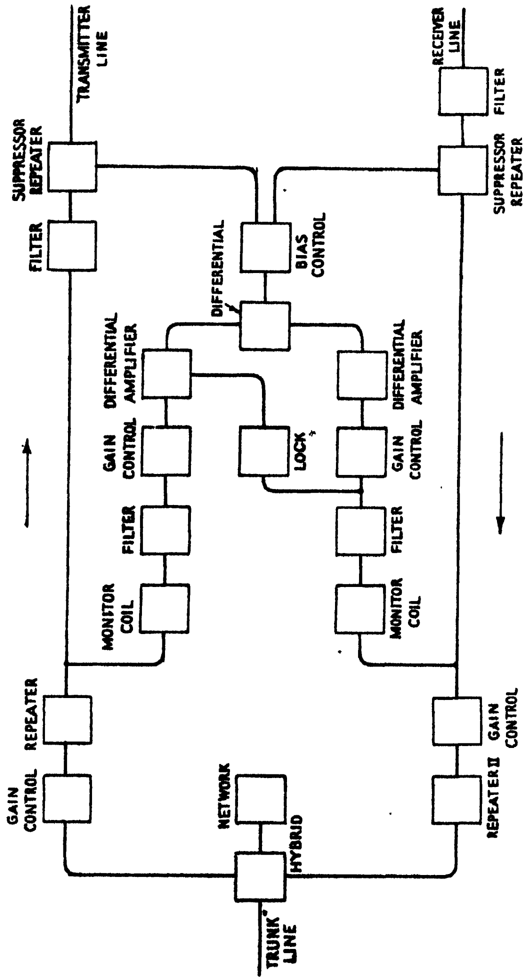


FIGURE 298.

sufficient time constant so that the suppressors are not changed over between connected syllables of the subscriber's speech.

Due to the incorporation of the differential arrangement the operation of the suppressors can be made satisfactory even when speech from the wireless receiver is at a considerably

higher level at the hybrid than the speech from the two-wire line at the same point.

Singing Suppressor—Marconi Type.[6] This equipment is in use at the various terminals in the British Empire with which the British P.O. conduct telephone services and, therefore, has to work in conjunction with the P.O. suppressor just described and in principle it is very similar, but a brief description of certain parts may be of interest. A schematic diagram is given in Fig. 298. If privacy equipment is in use an additional hybrid coil and suppressors will be required as in the P.O. type.

Valves having independently heated cathodes are employed although D.C. filament heating is used, because convenient methods of applying bias voltages are thereby rendered possible.

The differential principle is used, but the means of applying it are quite different, as will be seen from Fig. 299. One of the inputs has come from the receiving side and the other from the transmitting, each having come through an amplifier gain control and filter. Each input undergoes full-wave rectification so that uni-directional currents flow in the resistances OP, OQ, both currents flowing away from the centre point. If the inputs are equal (and the valves matched) there will be no voltage across PQ but any inequality of input produces a D.C. voltage which is applied at the grid of the first valves in the bias control circuit. If there is no output from the differential, V_1 is cut off and there is no P.D. across R_1 to bias the receiving suppressor, but the grid of V_2 is positive and hence there is a P.D. across R_2 to bias the transmitting suppressor. If the receiving input becomes greater than the transmitting input, V_1 is merely made more negative, but if the transmitting input becomes the greater V_1 becomes conducting and if the disparity is sufficient the receiving suppressor will be blocked and the transmitting suppressor opened. Owing to the condenser the change back to the receiving condition takes a time to be effected, which may be varied.

The above differential arrangement has an advantage over circuits in which a grid bias is balanced against a speech input. Because the inputs from the transmitting and receiving sides are of the same nature, the differential is linear, that is, a given difference between the two inputs always produces the same

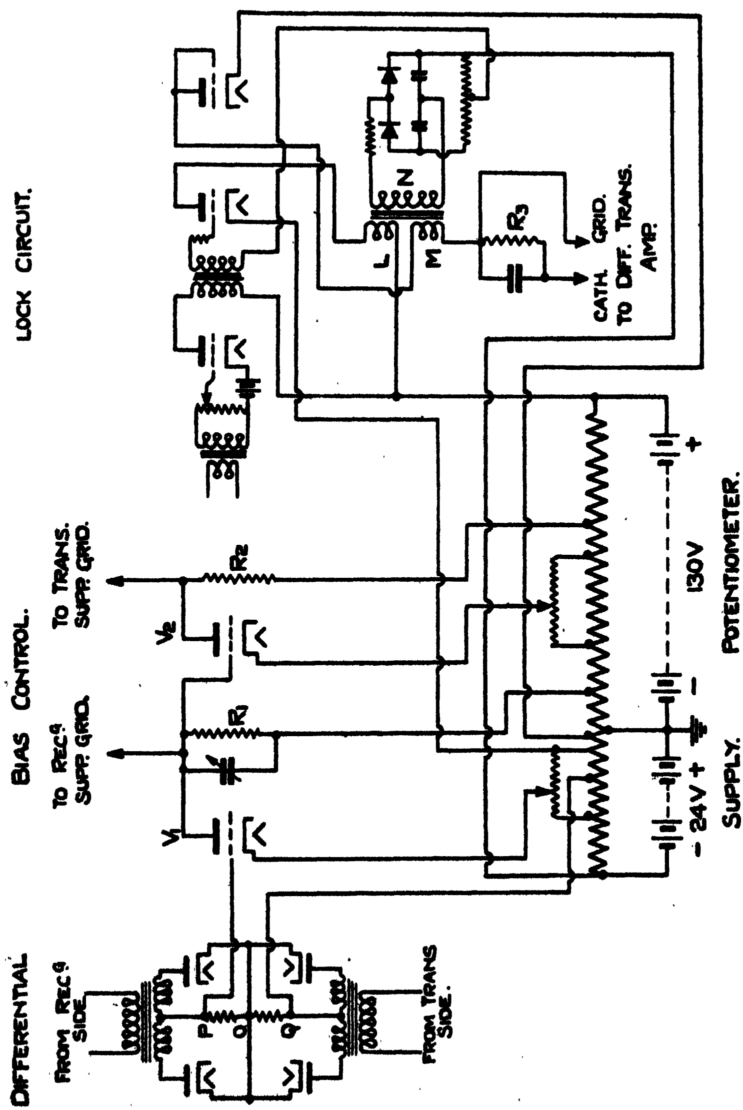

DIFFERENTIAL
FROM REC? SIDE
FROM TRANS SIDE
P
O
Q
BIAS CONTROL.
TO REC? SUPP. GRID.
TO TRANS. SUPP. GRID.
V1
V2
R1
R2
LOCK CIRCUIT.
L
M
N
R3
CATH.
GRID.
TO DIFF. TRANS. AMP.
- 24 V +
130V
SUPPLY.
POTENTIOMETER.

FIGURE 299.

P.D. across the output. In consequence of this, re-adjustment is not necessary if the noise level changes by a large amount.

A special feature of the equipment is the "Lock" Circuit, also shown in Fig. 299. This is so arranged that when the input from the receiving side exceeds a certain value the transmitting circuit is locked in the suppressed condition, irrespective of what voltages may be impressed on the differential from the transmitting side. The "Lock" is, therefore, so adjusted so that it does not work on received noise but does work on received noise plus received speech.

It will be seen that the first valve of the "Lock" Circuit is a straightforward amplifier taking its input from the receiving side whilst the second valve has a three-winding transformer in its anode circuit. Considering only the windings L and M it will be seen that as the input increases, the voltage applied to the diode increases, but because it is negatively biassed, the current in R_3 rises quickly after a certain input is reached. This effect is enhanced by the arrangement of metal rectifiers connected to the N winding which removes the bias on the second valve as the input increases, slowly at first, and then rapidly for larger inputs. The net effect is that at a certain value of input a large P.D. is produced across R_3 and this paralyses the differential transmitting amplifier. The condenser across R_3 ensures that the transmitting side remains suppressed until any echoes of the received speech returning from the two-wire circuit have been dissipated.

Some special methods of working will next be discussed before considering the question of privacy because some of these special methods give, incidentally, a measure of privacy.

Suppressed-Carrier Single Side-band System. The theory of this method of transmission has been fully dealt with in Chapter III, and the limitations discussed. The method has been employed on the long-wave trans-atlantic telephone circuit since its inception in order to economise in transmitter power and reduce the frequency band occupied. These considerations do not press so heavily on short wave telephone circuits whilst the difficulties are much greater due to the higher frequencies, but a suppressed carrier system is less liable to distortion when selective fading is present (see p. 115).

It will be evident that if the carrier fades deeply whilst the

sidebands remain at full strength the effect on the detector is the same as over-modulation and large second harmonics will be produced in the detector output. Any ordinary system of gain control depends upon the carrier and is, therefore, likely to increase the distortion when selective fading of the carrier occurs by causing the sidebands to be amplified unduly. Considerable improvement in quality might, therefore, be expected if the varying carrier were replaced by a steady carrier generated at the receiver. It is especially important to prevent the introduction of spurious frequencies when privacy systems are in use.

It is of interest to compare the signal/noise ratio obtained from an ordinary telephone transmission and from a suppressed carrier single sideband transmission. The basis of comparison and the assumptions made must be clearly understood or the figure arrived at is meaningless.

As a basis we will take it that the peak voltage in the final stage of the transmitter is the same for the two cases. When water-cooled valves are used, the limit of output will normally be the peak emission which the valve can give (see Chap. X), and hence by comparing on the basis of equal peak voltages we are really assuming that the final stage of the two transmitters will be the same.

The following assumptions are made :

(*a*) Sine wave modulation, the ordinary transmitter being 100% modulated.

(*b*) A distortionless transmitting path.

As an example, let the maximum value of the carrier voltage in the ordinary transmission be 10,000 volts, then the peak value with 100% modulation will be 20,000 volts and the maximum voltage of each sideband is 5,000. In the single sideband transmitter, however, the maximum value of the sideband will be 20,000 volts or four times that of each sideband in the ordinary transmission. If the transmission path is distortionless then the two sidebands of the ordinary transmission will arrive in the correct phase relationship and add their effects at the receiver detector. Hence the output voltage from the single sideband receiver will be twice that from the ordinary receiver, the power output being four times,

so that the gain is 10 $\log_{10} 4$, or 6 decibels approximately. The single sideband receiver need have only half the band width of the ordinary receiver and hence the noise energy picked up will be approximately halved. The signal/noise ratio in the single sideband receiver will, therefore, be 10 $\log_{10}$ 8, or 9 decibels above that in the ordinary receiver. This comparison is unable to take into account the effects of selective carrier fading and of phase displacement between the sidebands of the ordinary transmission.

In practice, the changing over of an existing channel to single sideband working would mean that the transmitter power could be greatly reduced during almost the whole working time.

The single sideband system has the serious disadvantage that it does not in itself provide a sufficient measure of secrecy and that the simpler form of privacy, that of speech inversion, is not suitable.

It is clear that the permissible variation in frequency between the suppressed carrier and the re-introduced carrier will be the same number of cycles whether long or short waves are used and hence the special difficulty in a short wave application of the method will be the very small percentage variation permissible in the re-introduced carrier. It is stated by Reeves[7] that a difference of 15 cycles between suppressed and re-introduced carrier is the maximum that can be tolerated on a commercial speech circuit. Such close agreement between independent oscillators would be exceedingly difficult to maintain and no such system has been tried out.

A number of alternatives present themselves, three of which have been the subject of extensive tests.

(*a*) The carrier and one sideband may be as completely suppressed as possible and a "pilot signal" applied at a low level as part of the modulation at the transmitter, this pilot signal being used to synchronise a local oscillator which re-introduces the carrier.

(*b*) One sideband may be suppressed and the carrier transmitted at a very low level. After separate treatment at the receiver the carrier may be re-introduced at the correct level.

(*c*) Transmission as in (*b*) but the carrier is used at the receiver to synchronise a local oscillator which re-introduces the carrier.

It should be noted that either the pilot signal of system *a*, or the low level carrier of system *b*, can be satisfactorily utilised at the receiver because they are practically single-frequency, unvarying signals and hence very narrow-band circuits can be used for them, thus allowing of a satisfactory signal-noise level ratio for these signals in spite of their low level at the receiver input. It should be further noted that where a local oscillator is to be held in synchronism it may be necessary to have circuits of large time constants in the

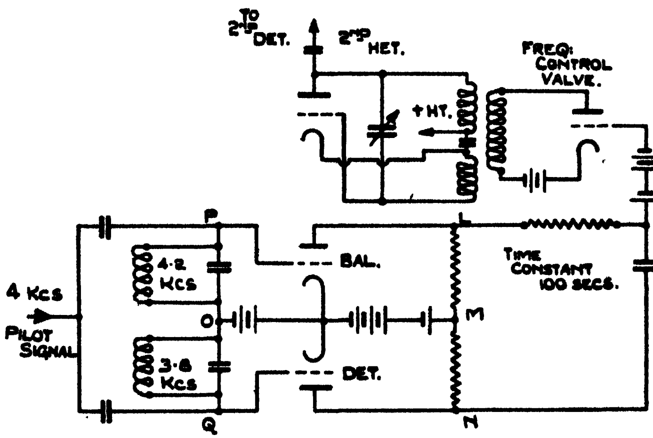


FIGURE 300.

synchronising arrangements to prevent the local oscillator drifting out of synchonism during periods of deep fading.

The receiver will always be of the superheterodyne type and one of the heterodynes will be automatically controlled, so that the beat frequency produced between it and the pilot (or low level carrier) is always correct, even if the transmitted frequency wanders, or if the heterodyne frequency tends to do so. The succeeding amplifiers may then have a band-width about equal to the band-width of the single sideband in order that the desired improvement in signal/noise ratio may be brought about, and it is also possible to use the narrow-band filters mentioned above for the synchronising signal.

An equipment based upon method (*a*) has been developed by the I. T. & T. Corp.[7] In this the pilot signal is 3·4 kc/s from the carrier and is, therefore, on the " edge " of the speech

band. The receiver is of the triple detection type and after the second detector, the pilot signal (now 23·4 kc/s.) is selected by very narrow filters. This signal is then made to beat with the output of a 27·4 kc/s. oscillator (which must remain constant to a high degree) and a 4 kc/s. output therefore obtained. This pilot signal also operates a gain control on the initial stages of the receiver, and in order to control the second heterodyne is fed into two sharply tuned circuits of equal decrement, one resonant at 4·2 kc/s. and the other at 3·8 kc/s. (see Fig. 300). If the frequency be exactly 4 kc/s. the voltage applied to each valve of the balanced detector will be the same and if the valves are matched there will be no P.D. between L and N. Under these circumstances the Frequency Control Valve has a certain resistance which has an effect upon the equivalent resistance of the oscillator circuit since it is coupled thereto.

If the frequency now rises above 4 kc/s. the voltage across OP will be greater than that across OQ and consequently the P.D. across LM will be greater than that across NM, so that the grid of the frequency control valve will become more negative and the resistance " thrown back " on to the oscillator circuit will be increased. This will slightly change the oscillator frequency and values can be adjusted so that the frequency variation which caused the 4 kc/s. signal to change is just compensated for and the correct band of frequencies obtained in the second intermediate frequency amplifier.

This arrangement would be perfectly satisfactory if the 4 kc/s. signal remained of constant strength, but in spite of the gain control applied to the first detector, changes will occur. Suppose the frequency is not exactly 4 kc/s., then there will be a difference in the P.D. between LM and MN as previously explained, but the actual magnitude of the difference will evidently depend upon the strength of the signal applied to the balanced detectors. Hence the frequency change produced in the second heterodyne would depend not only upon the frequency of the 4 kc/s. signal but upon its strength at the moment. Another gain control is, therefore, necessary worked off the point M and, therefore, responding to the sum of the balanced detector output, and applied to the pilot signal detectors.

Another equipment, developed by the Bell Telephone

Laboratories[8] and tested, in collaboration with the British Post Office, on a short wave trans-atlantic circuit, may be worked in accordance with either method (*b*) or (*c*) discussed above.

A schematic of the transmitter is shown in Fig. 301, the numbers in each section of the diagram indicating the frequencies (in kilocycles per sec.) of the currents traversing that section in a typical equipment. It will be seen that the carrier and one sideband are first suppressed by a balanced modulator (see p. 261) and special crystal filter arrangement at a comparatively low frequency and a variable amount of carrier is

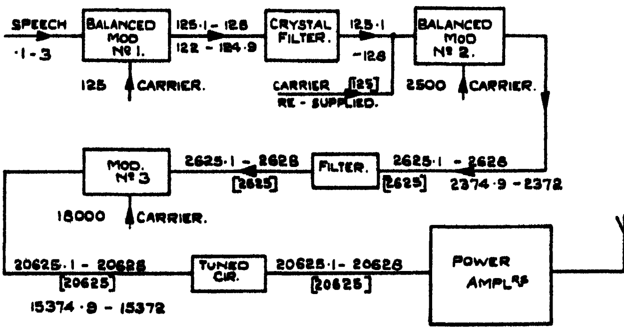


FIGURE 301.

then re-introduced. Only the circuits from the third modulator onwards need altering if the transmitted frequency is to be altered. The high-power stages of the transmitter are quite normal.

The receiver is of the double-detection type and is shown schematically in Fig. 302. The frequency of the first heterodyne is automatically adjusted so that the carrier frequency in the intermediate frequency amplifier is matched to the suppression band of a crystal band-suppression filter (shown in Unit A, Fig. 303, to be described later). The intermediate frequency amplifier can, therefore, have a band-width limited practically to the width of one sideband.

The carrier is selected by a crystal band-pass filter (shown in Fig. 302) amplified and passed through a series of amplitude-smoothing arrangements and may then be applied to the second detector as the necessary re-introduced carrier.

It may at first be thought that there will be no improvement in performance over a normal system when selective fading is present, since the original carrier is being used in the final detector, but it should be realised that, because the carrier is available by itself at the receiver, it may be passed through amplitude smoothing circuits which could not deal with the whole signal because they would then smooth out the modulation. These smoothing circuits take the form of overloaded amplifiers.

As an alternative arrangement—and a more satisfactory

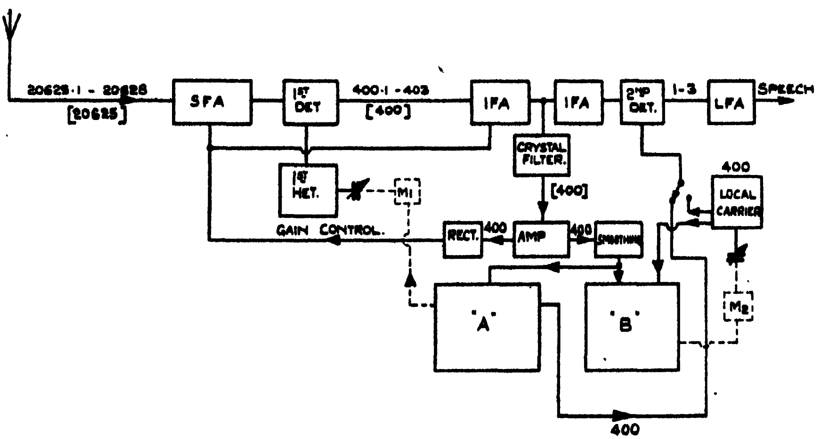


FIGURE 302.

one when the field strength is very low—the original carrier may be used to synchronise a local oscillator which is applied to the second detector.

Quite original and ingenious arrangements (shown in Fig. 303) are used to adjust automatically the first heterodyne, and will now be briefly described. The carrier is applied to the input of the push-pull pair shown in Fig. 303A, and also to the grid of a screened grid valve, the output of which contains a crystal band-suppression filter. Coupled to the parallel input of the push-pull pair is a 60 cycle A.C. supply which also feeds the voltage coil of an induction type watt-hour meter, the current coil of which is in the output of the push-pull pair. If the carrier emerging from the intermediate frequency amplifier is exactly the same as that of the crystal

filter then there is no high frequency voltage E and equal voltages E_1 and E_2 are impressed on each grid. The 60 cycle voltage is small compared with the high frequency voltages. If now the carrier begins to drift away from the

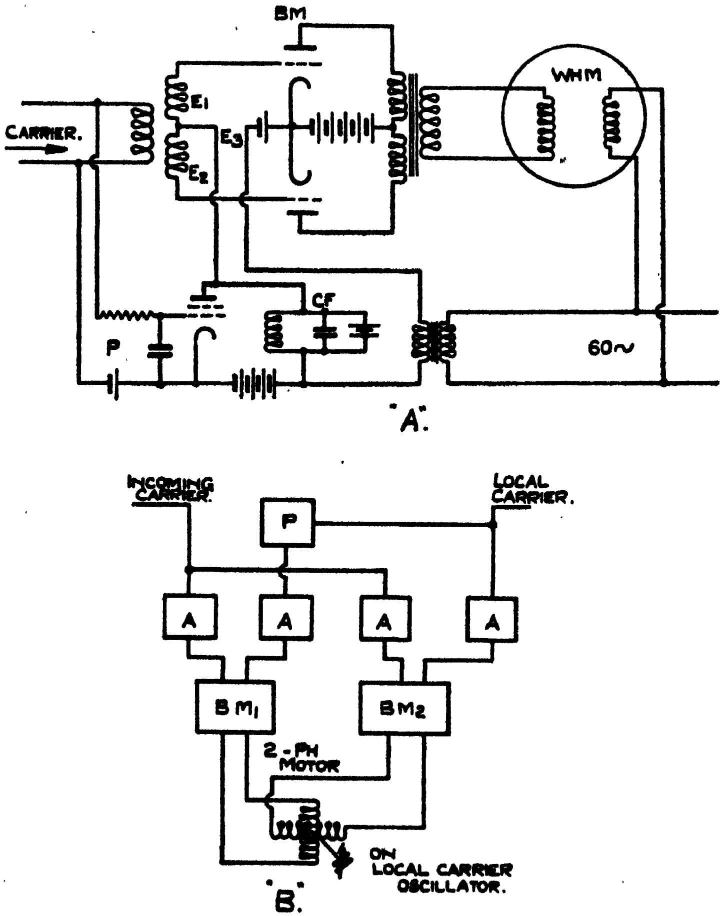


FIGURE 303.

crystal frequency the voltage E_3 appears, its magnitude and phase depending upon the direction and extent of the frequency drift. The resultant voltages to the grids are, therefore, $E_1 + E_3$ and $E_2 + E_3$ (vector summations). One of these resultants will be large and the other small (since E_1 and E_2 are opposite in

phase) and the valve which has the greater high frequency voltage applied to it will have the greater mutual conductance (since the valves are initially biassed to the bottom end). More 60 cycle current will, therefore, flow through one valve than through the other, and hence the current through the current coil of the watt-hour meter will depend upon the magnitude and direction of the frequency drift. If the carrier frequency is higher than that of the crystal, the disc of the meter will run in one direction, whilst if the carrier frequency is lower the rotation will be reversed. The disc of the meter drives (through gearing) the moving vane of a very small condenser in the first heterodyne and thereby compensates for the frequency drift.

When it is desired to use a local oscillator at the second detector an additional synchronising arrangement is required to synchronise the local oscillator to the incoming carrier because the above arrangement, since it merely refers the carrier to a crystal filter, would not be an accurate enough control for a re-introduced carrier.

The local carrier is supplied to two separate push-pull detector circuits, one directly and the other through a 90° phase-shifting circuit. The incoming carrier is supplied to the same two detector circuits. Each detector circuit will, therefore, give an output at the beat frequency between the local oscillator and the incoming carrier and the two outputs will be quadrature. The outputs are supplied to a special type of two-phase motor, this driving the moving vane on a condenser in the local oscillator circuit. The direction of rotation of the motor will depend upon the direction of frequency drift, and when this has been compensated for the motor will stop since the frequency supplied to it will be zero.

The incoming carrier is used (before smoothing) to provide gain control on the high frequency amplifier and first detector.

The results of tests with both the I. T. & T. Corp. and Bell equipment showed that the gain in signal/noise ratio was of the order predicted by theory, and articulation tests showed a considerable improvement in the circuit apparently due to absence of selective fading effects.

Dual-Channel System. An arrangement is in use on the short wave trans-atlantic telephone circuits, whereby two sidebands are transmitted but the lower side-band is carrying

one telephone channel and the upper side-band a second channel. Referring to Fig. 301 it will be seen that it is possible to have a duplicate No. 1 balanced modulator, modulated by another speech channel. If the crystal filter following this selects the lower side-band, then the outputs of both filters can be applied to No. 2 balanced modulator and the rest of the equipment is much as before, except that twice the band-width must be allowed for.

In the actual equipment, the carriers supplied to the two No. 1 modulators differ by about 3 kc/s and in the final transmission, therefore, one channel is displaced by this amount from the carrier. This is necessary because the low-level carrier which is transmitted beats with both side-bands and would produce " cross talk " between the channels. When one channel is displaced the inter-modulation products fall into the vacant " space " between carrier and displaced channel.

At the receiver, both channels are kept together as far as the output of the 1st I.F. amplifier. Two 2nd-frequency changers are provided and the two 2nd I.F. amplifiers are sufficiently selective to each pass only one side-band.

The "Ray-Diversity" Receiver. In Chapter VIII we mentioned the Multiple-Unit Steerable Array and it is now necessary to consider the type of receiver to be used with it. As the receivers used are somewhat complex, space only permits of us giving the barest outline and the paper by Gill[12] should be consulted for further information. Referring to Fig. 304 which shows a greatly simplified schematic diagram it will be seen that the sixteen rhombic arrays each have their own R.F. and 1st I.F. circuit. At the second frequency-changer arrangements are made to correct the phase of the inputs from the 16 aerials so that they may be combined. This is done by supplying local-oscillator voltages differing in phase. The phase-differences are obtained by an ingenious arrangement of artificial line and auxiliary oscillator (detail not shown in Fig. 304), the adjustment being the frequency of the latter.

Four separate phase-adjusting circuits are available so that four separate zenithal polar-diagrams can be obtained at the same time. Hence from the second frequency-changer there are four separate circuits A, B, C, and M, each corresponding to a certain (adjustable) reception angle. Three of these

circuits, after further amplification and rectification, are combined at audio frequency as indicated.

Thus if the energy is reaching the receiving aerials by a number of rays, the three best of these can be selected and the energy finally added. Delay networks can be inserted at audio frequency to allow for the fact that the ray paths may be

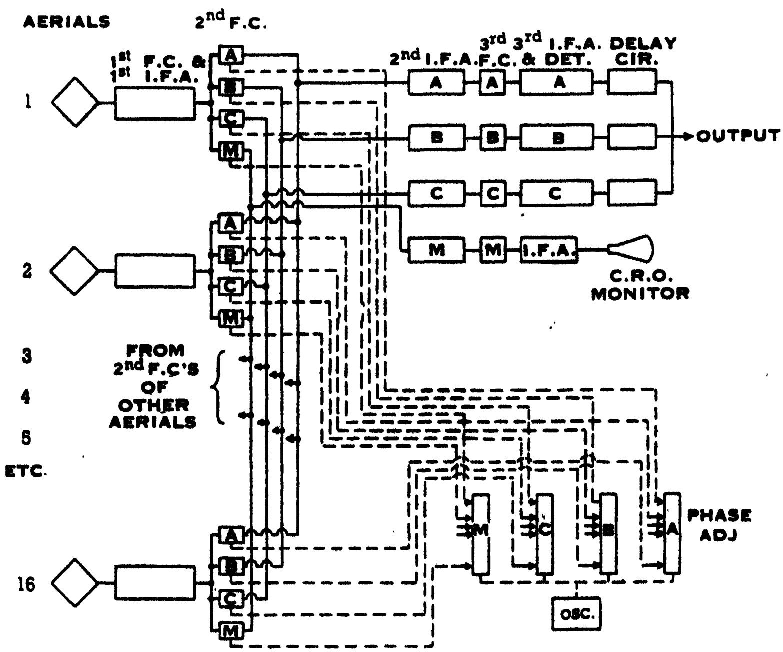


FIGURE 304.

of different lengths. If there are less than three rays the unwanted circuits may be cut off.

The fourth path *M* through the receiver is the monitoring circuit which terminates at a cathode-ray oscillograph. The oscillator phase of this path is being continually varied so that its zenithal angle of maximum reception swings over a range of values periodically, and an inspection of the oscillograph screen enables one to see at a glance at what angles useful rays are coming in and to adjust the *A*, *B*, and *C* paths accordingly.

The actual receiver is for dual-channel, single side-band working and a number of receivers could be connected to the same array system.

Multi-Channel System. In modern line-telephony it is a common practice to use the carrier system in order to work a number of channels over one line circuit. Thus, in a standard equipment, 12 carriers, spaced between 12 and 60 kc/s, are each modulated by a separate speech channel. These channels will normally all carry speech in the same direction and another pair of wires carry the return channels, that is, the system is worked on a four-wire basis.

In some cases where ultra-short waves are employed by the British Post Office for linking together telephone lines which are working on this principle, all the lower-frequency carriers are used to modulate one very high-frequency carrier. That is, one wireless circuit carries all the channels.

A disadvantage of such a system is that the mean percentage modulation applied to the wireless transmitter by each carrier channel must be very low in order that should a number of channels rise to peak modulation at the same instant, the wireless transmitter is not overloaded.

A simpler system for obtaining only two channels over a line is to use the original speech frequencies for one channel and to use a 5 kc/s carrier for a second channel—the upper sideband being selected. Hence the whole frequency band extends to 10 kc/s. In a number of the P.O. ultra-short wave links, these two channels modulate a single ultra-H.F. carrier.

On these ultra-short wave circuits, the wireless transmitters and receivers are so designed that they can be left unattended and controlled from a distant telephone exchange.

Quiescent Carrier Systems. It is evident that a considerable saving of power can be made if a telephone transmitter has its carrier emission stopped whenever speech is not actually being transmitted. Actual measurements taken by the Post Office indicate that during an ordinary conversation each transmitter is only transmitting speech for about 13% of the time and there will be many pauses between conversations when neither transmitter is actually in use.

In addition to the economy effected, "eaves-dropping"

becomes more difficult since the carrier is only transmitted for brief periods.

When one end of the telephone channel is on shipboard it is almost essential for the transmitter on the ship to work on the quiescent-carrier principle because of the very small separation possible between transmitter and receiver. Comparatively heavy currents are induced in rigging adjacent to a short wave aerial and the vibration of the rigging causes loud crackling noises to be produced in nearby receivers.

One method of achieving carrier suppression is by a modification of the absorber keying circuit described on page 425. The

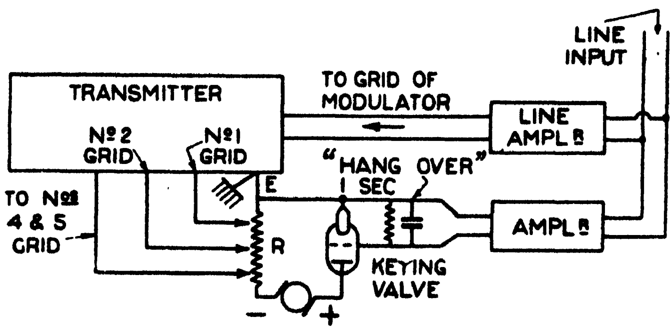


Figure 305.

incoming speech is fed into the input of a control amplifier, Fig. 305, as well as into the line amplifier which modulates the transmitter. The control amplifier is followed by an anode bend rectifier (included in the amplifier in Fig. 305), the D.C. output voltage of which is applied to the grid of a valve which takes the place of the keying relay.

With no speech, since the rectifier valve of the control amplifier is then passing no feed, the grid of the keying valve is at positive potential and a heavy feed flows through the keying resistance, causing the fourth stage to be cut off, and thus preventing radiation of the carrier. When speech is applied the rectifier takes feed, the keying valve grid is backed off negative, the current through the resistance is stopped, and the circuit drives in the normal manner.

It will be realised that it is not desirable such a circuit should cut off the carrier in the intervals between syllables of speech,

and to prevent this a time constant circuit (shown in Fig. 305) is provided whose constants can be arranged such that only for "pauses" will the carrier be suppressed.

A difficulty is introduced at the receiver by the employment of quiescent-carrier working at the transmitter. The normal type of automatic gain control depends upon the carrier for its operation. If, therefore, the carrier fails, the gain control will push up the amplification to the extreme limit, thereby raising the noise at the receiver output to a very high level. Since no speech is being received at the moment, this might not appear to matter, but this noise passing along to the terminal equipment is liable to lock the suppressors in the "receive" connections and prevent transmission.

One solution is the use of a similar arrangement to that employed in some broadcast receivers in order to prevent an irritating noise from the loudspeaker when tuning between stations. All the stages of the receiver following the gain control are kept paralysed unless a signal of a certain minimum strength is picked up. This is more difficult to do on a commercial short-wave receiver because the minimum signal strength it is required to receive is so much smaller than in the case of a broadcast receiver.

The Compandor System.[9] The name of this system is a compound of compressor and expandor because at the transmitting end of the circuit the amplitude range of the signal is compressed, whilst at the receiver it is expanded again. The object is to improve the signal/noise ratio and, as this is poorest when long waves are used, the method is only used at present on the long wave trans-atlantic circuit but is included here for completeness.

In order to obtain the best signal/noise ratio it is evidently necessary that the transmitter should always be fully modulated. The strength of speech arriving at the terminal equipment from the two-wire circuit naturally varies greatly with the type of connection and peculiarities of subscriber. It is estimated that on an ordinary telephone system there is as much as 70 dbs. difference of level between a consonant as rendered by a soft-voiced speaker and a vowel spoken by a loud-voiced one.

Manual adjustment at the terminal equipment is always

resorted to in order to adjust for widely differing subscribers, and this reduces the amplitude range to something like 30 dbs. but cannot, of course, allow for sudden changes or the very different energy levels of different syllables. It follows that on a circuit where the signal/noise level is poor, if the highest level portions of speech fully modulate the transmitter the lowest level (but very significant) portions may fall below the noise level.

In the Compander System the difference in level is reduced to about 15 dbs. by the compressor and then expanded at the receiver, usually to the original amplitude range.

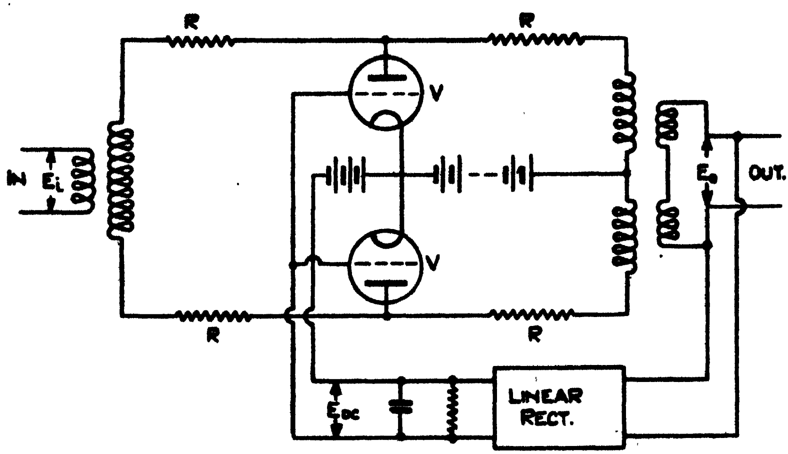


FIGURE 306.

The circuit of the compressor is shown in Fig. 306. The valves VV are used as variable resistances and it will be seen that if their resistance is decreased, the voltage available at the output for a given input will be decreased. The valves are set on the curved portions of their characteristics so that R_B, the anode-filament resistance, is inversely proportional to E_{DC}.

If the output voltage at a given time is E_O, then this may be written $K_1 E_i R_B$ where E_i is the input voltage whilst K_1 is a constant dependent upon the transformer and resistances $R R R R$. E_{DC}, the value of the D.C. output from the linear rectifier, is given by $K_2 E_O$ and

$$R_B = \frac{1}{K_3 E_{DC}} = \frac{1}{K_3 K_2 E_O}$$

Substituting this value for R_B in the first equation for E_O gives

$$E_O^2 = \frac{K_1 E_i}{K_3 K_2} \text{ or } E_O = K\sqrt{E_i}$$

Hence, if the level of E_i varies from time to time by 30 dbs., the level of E_O will vary by only 15 dbs.

The resistance-condenser combination in the rectifier output has a small time-constant so that the compressor works at about syllabic frequency.

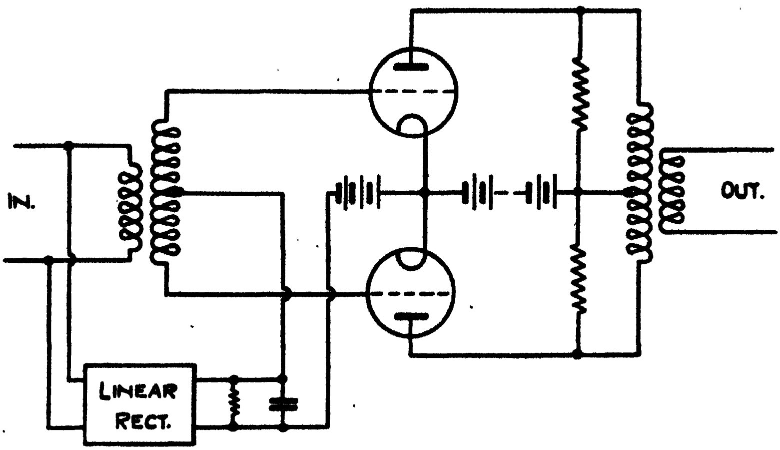


FIGURE 307.

The expander circuit used at the receiving end is shown in Fig. 307, and it will be seen that the valves are biassed to the curved portions of their characteristics and that a portion of the input is rectified and reduces this bias. The amplification of the received signal will, therefore, be approximately proportional to the amplitude of the input and hence the amplitude compression of the transmitted speech will be removed.

It will occur to the reader that the use of the curved portions of the valve characteristics is likely to result in the production of unwanted even harmonics. This is prevented by the use of push-pull circuits.

Privacy Systems—A General Discussion.[10] Although nothing seems easier than the accidental production of unintelligible speech, yet when we wish to render speech unintelligible

to all except one special receiver the matter becomes difficult, and to give successful results requires an elaboration of apparatus.

The special methods of working which have been discussed in the previous section do provide in themselves some measure of privacy if undistorted speech is applied to the transmitter, as any one equipped with simple receiving apparatus would need to have very fine adjustments if they wished to pick up the transmission with any degree of intelligibility. These methods, however, can hardly be regarded as secret systems in themselves.

The privacy systems most used are voice frequency systems situated at the terminals of the circuit and are, in fact, equally applicable to line working if desired.

By a suitable arrangement speech may be "inverted," that is low frequency components of the speech become high, and vice versa.

An alternative is to separate the speech frequencies into bands by means of sharp filters and then change the frequencies so that the various bands are transposed. This is termed "scrambling."

Privacy systems have also been developed in which a varying delay is introduced so that the syllables of the speech are transmitted in the wrong time order and are shifted back at the receiver by a synchronised device. These systems have the disadvantage that they increase the transmission time of a circuit in which they are used.

These methods require elaborate apparatus, but the same apparatus can be used for both directions of speech, by a suitable arrangement. If reference is made to the complete circuit of Fig. 297 which shows the Post Office equipment, it will be seen that this is accomplished in the following way.

It will be noticed that in addition to the privacy equipment, additional singing suppressors RS_2, TS_2 are necessary, and also a second hybrid. The control for the receiving suppressor RS_2 is linked in parallel with the first receiving suppressor RS_1, and the control for TS_2 is linked in parallel with TS_1. This means that when RS_1 is clear so is RS_2, and TS_1 and TS_2 will be blocked. This is the condition for incoming speech.

Thus speech from the wireless receiver passes through RS_2 and from the hybrid through the privacy equipment, which inverts it, and into the terminal in the ordinary way. The connection from the output of the privacy equipment to the suppressor TS_2 can do nothing, because TS_2 is blocked.

Considering the condition when transmitted speech is operating the gear, both suppressors RS_1, RS_2 are closed and TS_1, TS_2 are open. This means transmitted speech passes up through the hybrid to the privacy equipment, is inverted, and the inverted speech passes through the transmitting suppressor TS_2 to operate the main transmitter.

Privacy by Inversion of the Whole Speech Band. If a speech wave having frequencies extending from 250 to 2,750 cycles be used to modulate a 3,000 cycle carrier, then two sidebands, 250 to 2,750 cycles and 3,250 to 5,750 cycles are formed. If now we select the lower sideband it will be observed that the speech is "inverted," for it is the 2,750 cycle component of the speech which produces the 250 cycle frequency, whilst the 250 speech component produces the 2,750 cycle frequency. Notice that if we had inverted speech and used this to modulate the same carrier, then selection of the lower sideband as before would lead to straight speech. Thus a single piece of inverter apparatus could be used for the dual purpose of inverting or re-inverting speech.

Privacy systems utilising this principle have been developed, and are in use, but the simple method of inversion indicated above is not usually adopted at the moment for the higher grade apparatus because of the difficulty of separating the inverted speech from the original speech and from the 3,000 cycle carrier. Instead, a double modulation system is used, for example, the speech may initially modulate a carrier of 13,000 cycles per second, and if the speech frequency is represented as S cycles per second, resultant frequencies of $13 - S$, 13, and $13 + S$ kilocycles per second will be obtained. Of these $13 - S$ and 13 are suppressed, and the remaining sideband, $13 + S$ kilocycles is used to modulate an oscillator having a frequency exactly three kilocycles per second in excess of the first oscillator, that is, 16 kilocycles.

Thus the output from the second modulated oscillator will comprise frequencies of $3 - S$, 16, and $29 + S$ kilocycles per

second; the two latter are suppressed, and the remaining sideband forms inverted speech. This is again seen to be a low frequency band, and is used to modulate the outgoing high frequency system in the ordinary way.

By such a double modulation system the original speech can be completely cleared from the system, and the design of filters for separating the various sidebands becomes more easily possible.

Such a method of inversion is not entirely secret. If an ordinary receiver of the self-oscillating detector type be adjusted to oscillate 3,000 cycles below the incoming carrier, owing to the selectivity of the receiver circuit, the lower sideband terms will predominate and the interference beats with the receiver carrier will cause inversion, although, of course, a strong 3,000 cycle tone will also be heard.

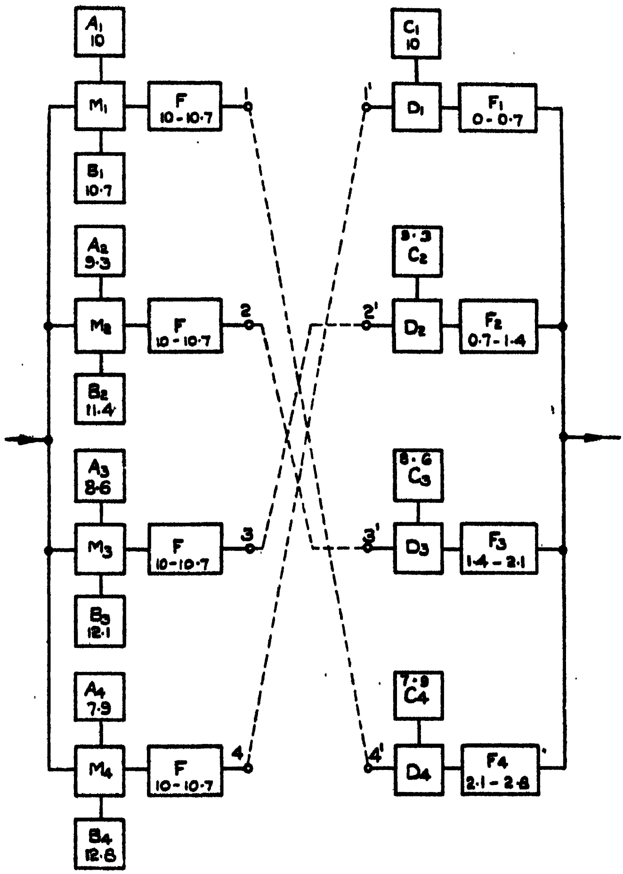
Interception in this way may be prevented by wobbling the carrier frequency of the wireless transmitter about $\pm$ 500 cycles at a very low rate, usually two or three times a second. This prevents re-inversion by the method just mentioned, but makes no difference to reception by the proper method wherein the carrier is eliminated at the detector as usual, and privacy equipment used to re-invert the speech.

It will be appreciated that the above method of privacy is of no use if applied to a single sideband transmission because merely shifting the re-introduced carrier by (in our example) 3,000 cycles, will make the speech normal. Additional privacy by wobbling the carrier frequency is no longer possible unless a synchronous wobble of suppressed and re-introduced carrier can be managed.

Privacy by Speech "Scrambling." This is best illustrated by an example, shown schematically in Fig. 308. The speech (assumed to have a band-width of 0 — 2,800 cycles) is applied in parallel to four modulators, M_1, M_2, M_3, M_4, each of which combines the speech with the output of *either* its associated A *or* B oscillator. The output of each oscillator is applied to a filter, F, which in every case has a pass band of 10,000–10,700 cycles.

Suppose A_1 is in use, then sidebands 10,000–12,800 and 10,000–7,200 cycles will be produced at M_1, but only the frequencies 10,000–10,700 cycles will pass F. The output from

1 is due, therefore, to those components of the input speech which had frequencies between 0 and 700 cycles and these components are "upright," that is, the lower input frequencies are still the lower frequencies.



NUMBERS ARE KILOCYCLES PER SEC.

FIGURE 308.

Now suppose B_2 is in use. Sidebands of 11,400–8,600 cycles and 11,400–14,200 cycles will be produced at M_2 and 10,000–10,700 cycles only will be passed by F so that the output at 2 is due to components of the speech having frequencies between 700 and 1,400 cycles, but because the lower sideband has been selected, these components are "inverted."

In the same way the remaining two circuits will deal with the rest of the speech components, the outputs being " upright " or " inverted " according to whether the A or B oscillator is in use, and all four outputs, each containing different components of the speech will have frequencies between 10,000 and 10,700 cycles. On this account it is possible to " jumper " any output of this stage of the privacy equipment to any input of the next stage, but in order that an identical equipment at the receiver shall re-convert to plain speech, it is essential that the " jumpering " be complementary to the two equipments, that is, if 1 is connected to 4′ in one equipment, then 4 must be connected to 1′ on the other equipment, and so on. The speech must, therefore, be split into an even number of bands.

The second portion of the equipment consists of four detectors, D_1, D_2, D_3, D_4, each having an associated oscillator C_1, C_2, C_3 and C_4, the frequencies of which are arranged (as shown in Fig. 308) to produce speech frequency bands between 0 and 2,800 cycles when the inputs to the detectors are all 10,000–10,700 cycles per second.

If the arrangement shown be adopted, the output from 1 will comprise frequencies 2,100–2,800 and a series of distortion terms. The band 2100–2,800 is selected by F_4 and hence these frequencies appear in the output due to components of the input between 0 and 700 cycles. Similarly, the other speech frequencies may be followed through the equipment, and it will be seen that the output to the line occupies the frequency band 0–2,800 cycles, but that the relationship between the original and output frequencies is most disordered. An " eavesdropping " receiver requires not only elaborate equipment, but also a knowledge of the " jumpering " and this can be frequently changed.

It is necessary to assure ourselves that an identical equipment will give us clear speech at the output if the distorted speech is fed into the input. Suppose we trace the original speech frequencies between 700–1,400 cycles. These left the transmitting privacy equipment as frequencies from 1,400–2,100 cycles, the band being " inverted." This band of frequencies on reaching the receiving privacy equipment is combined in M_2 and B_2 to form a sideband of 10,000–10,700 cycles which is re-inverted, but still contains the frequency band 1,400–2,100.

This is combined at D_2 with the output of C_2 and the frequencies 1,400–2,100 cycles are changed to 700–1,400, freed of distortion by the filter F_2 and pass to line as the original band.

It should be pointed out that this particular method is not in general use.

Associated Measuring Apparatus. (See Fig. 309.) In order that wireless telephone channels may be efficiently worked it is highly desirable to include in the terminal equipment certain measuring apparatus, so that the performance of the circuit from day to day can be expressed in a more or less quantitative way and any preventable deterioration corrected

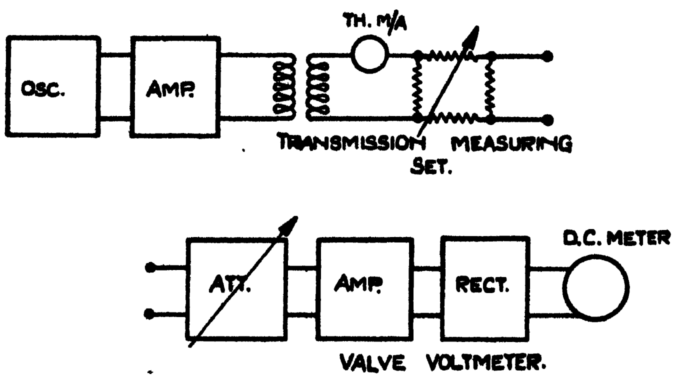


FIGURE 309.

for. The use of such apparatus also assists greatly in the localisation of defects or breakdown.

An oscillator will be necessary as a source of power when taking measurements and will be used for the preliminary "lining up" of the whole circuit.

The oscillator will usually be of the heterodyne type, that is, the output note is produced by combining and rectifying the outputs of two high frequency oscillators, one of these being of large amplitude and the other small. By this means an output nearly free from harmonics is obtained and a wide range of output frequencies may be obtained by a small adjustment of a variable condenser of moderate size in one oscillator.

The oscillator output will be fed through an amplifier having adjustable gain into the transmission measuring set. The first portion of this contains a thermal milliammeter and a

network so that this portion can be adjusted to be the equivalent of an alternator having an internal resistance of 600 ohms (and no reactance) and giving an output of 1 milliwatt. Following this is a network of resistances which may be adjusted so that the final output level can be reduced, usually in 1 db. steps, the resistance remaining at 600 ohms.

The next part of the measuring apparatus to be considered is the valve voltmeter. The input to this can be arranged to be of high impedance so that it may be bridged across a circuit without disturbing it. A variable, calibrated attenuator is followed firstly by an amplifier, then by a rectifier, the output of which is shown on a moving coil milliammeter.

These components may be used together in a variety of ways. The valve voltmeter is calibrated by producing a known power output from the transmission measuring set and applying this direct to the valve voltmeter. The attenuator is adjusted to give a certain reading on the rectifier milliammeter. If now the loss (or gain) in decibels occurring in a certain portion of the terminal equipment is to be measured, the output of the transmission measuring set will be applied to the valve voltmeter via the apparatus under test and the attenuator adjusted till the rectifier milliammeter reads the same as during calibration. The difference in attenuator settings for test and calibration evidently gives the loss (or gain) in decibels occurring in the apparatus under test.

If the gain (or loss) in the whole wireless telephone circuit is to be measured then the transmission measuring set at one terminal point will be adjusted to give a known output and this applied at the desired point in the terminal equipment. The received signal will be measured by the valve voltmeter at the other terminal point (after calibration by means of the transmission measuring set at its own terminal).

We have so far only considered measurements conducted when a pure tone is in use but evidently measurements of levels when speech is passing along the circuit will be necessary and it is also desirable we should be able to measure the noise level.

The noise level may be measured on the valve voltmeter in the same way that pure tones are measured, except that the instrument pointer will be rather unsteady due to fluctuations in noise intensity.

In order to monitor speech levels each terminal equipment will include one or more volume indicators. These may consist of a single-stage amplifier having a high input impedance followed by a copper-oxide rectifier and moving coil milliammeter. The milliammeter is scaled in decibels below or above a reference level, the calibration having been obtained with the aid of an oscillator and transmission measuring set, or similar equipment. The volume indicators may be plugged across various points in the incoming and outgoing circuits and, though calibrated by means of a pure tone, give useful direct indications of speech and noise levels. They will not be so accurate as the valve voltmeter previously described since they depend upon the permanence of an amplifier.

The measurement of speech level cannot be made in the same precisely defined units as the pure tone measurements because the speech varies so rapidly in intensity. All that can be done is to adopt some definite, even if quite arbitrary method of measurement so that tests taken at different terminals which work together may be comparable.

Operation of a Circuit.[11] It may be of interest to outline the methods of operation and control adopted by the British P.O. at their London terminal.

Each set of terminal equipment is in charge of a technical operator who, in association with his colleague at the other end is in control of the channel.

The actual connection of subscribers, etc., is carried out by traffic operators in the Radio Section of the International Telephone Exchange which adjoins the terminal. The switchboard equipment and method of operating is very similar to that of an ordinary trunk exchange except that the calls are supervised throughout in order that extra time may be allowed if circuit conditions become bad.

Order wires extend from the technical operator's position to the traffic operator and to the wireless transmitting and receiving stations.

About half an hour before a service is scheduled to be available, the transmitters, etc., are started up and the two technical operators get into touch with each other, the preliminary calling being usually by tone telegraphy. The circuit is then "lined up" by passing a pure tone from the

oscillators at each end and the reading of the volume indicator across the outgoing line when the transmitter is fully modulated noted. The noise level is also measured and the singing suppressor adjusted accordingly.

The traffic operators are now notified that the circuit is available by an arrangement of coloured lights which show whether the circuit is fully available or cannot be satisfactorily extended by long line circuits. This latter condition may arise if the signal/noise ratio over the wireless circuit is unusually poor.

The technical operator can light up calling lamps on the traffic operator's position if the distant terminal is calling. This is necessary since the traffic operators are not attached to any particular channel, whilst the technical operators are always on duty on their particular channel.

As different subscribers are connected to the circuit the technical operator will adjust the gain of the transmitting amplifier to keep the wireless transmitter fully modulated at the peaks of speech. This is very important since if the transmitter modulation falls off seriously the noise level at the receiver will become poor.

SELECTED REFERENCES

1. *Trans-Atlantic Telephony.* Beer and Evans. *P.O.E.E.J.*, Vol. 20; April, 1927.
2. *Overseas Extensions to Wire Telephone Networks.* Espenschied and Wilson. *P.I.R.E.*, Vol. 19; February, 1931.
   (Gives extensive bibliography of papers on wireless telephony.)
3. *Electrical Communication.* February, 1930.
   (Special issue devoted to short wave wireless telephony.)
4. *Radio Telephony Terminal.* Scarborough. *P.O.E.E.J.*, Vol. 24; April, 1931.
   (Gives description of Relay-Type Singing Suppressor.)
5. *Methods for Measuring Interfering Noises.* Espenschied. *P.I.R.E.*, Vol. 19; November, 1931.
6. *Terminal Equipments.* Murphy. *Mar. Rev.*, Nos. 48 and 49; 1934.
7. *Single Sideband System Applied to Short-Wave Telephone Links.* Reeves. *J.I.E.E.*, Vol. 73; September, 1933.

8. *A Single Side-Band Short-Wave System for Transatlantic Telephony*. Polkinghorn and Schlaack. *P.I.R.E.*, Vol. 23 ; July, 1935.
9. *The Compandor*. Mathes and Wright. *Bell Sys. Tech. Jour.*, Vol. 13 ; July, 1934.
10. *Privacy Systems for Radio Telephony*. A. J. Gill. *P.O.E.E.J.*, Vol. 26 ; October, 1933.
11. *British P.O. International Exchange*. Birch and Hartwell. *P.O.E.E.J.*, Vol. 27 ; July, 1934.
12. *Chairman's Address*. Wireless Section. Gill. *J.I.E.E.*, Vol. 84 ; 1939.

CHAPTER XVII

COMMERCIAL TRANSMITTERS AND CIRCUITS

THE design of a short wave transmitter varies very considerably, depending upon whether it is for general purpose work or point-to-point commercial communication. Generally speaking, the former will be of small power, it need not produce an extremely constant frequency, but it will require to cover a wide range of wavelengths. It must be provided with means for rapid "changing over" and will have to be designed to work on C.W., tone-modulated C.W. (i.e. I.C.W.), or telephony. Such a set will be suitable for ship installations, and services where very high speeds are not required but flexibility is important, and low prime and running costs a consideration.

It is with the large transmitter that design has undergone considerable change. This is largely because of the higher performance that is called for, and the fact that there is a growing demand for greatly increased powers on short waves. Whereas a few years ago, 50 kW was considered to be an adequate power for a short wave transmitter for any purpose, powers now are ranging to above 100 kW. It may be of interest therefore to indicate briefly the type of specification that such a transmitter has to meet, before giving a brief description of a high power transmitter, similar to that installed in the Empire Broadcasting Station at Daventry.

The B.B.C. specification required an output of 100 kW unmodulated, at wavelengths between 30 and 80 metres, and 75 kW at 13 metres, rising to 100 kW at 30 metres, and that a quick change to any of four pre-determined frequencies should be possible. Crystal oscillators maintaining the frequency to within $\pm$ 1 part in 100,000 were to be provided for each of eleven operating frequencies to be specified. In addition, a resonant circuit valve oscillator was to be provided to enable other frequencies to be obtained, if desired, the con-

stancy required for this being within $\pm$ 1 part in 25,000. The scintillation during modulation to be less than one cycle, with 80% modulation.

It was further specified that the R.M.S. sum of the harmonic content on any frequency of modulation should not exceed 4% of the voltage of the fundamental frequency when the depth of modulation was 90%. The frequency response was to be constant to within $\pm$ 2 db. from 50 to 8,000 c/s.

Marconi 100 kW Short Wave Broadcast Transmitter. Diagrams of connections for a transmitter to meet this specification are shown in Figs. 310 and 311, and constructional details can be observed by reference to the views of the transmitter shown in Figs. 312 and 313.

Crystal Unit. The crystal maintaining circuit, Fig. 310, which is of the Pierce type but employing a pentode valve, the valves, and eleven crystals, one for each of the eleven "spot" waves, are all housed within a common temperature controlled chamber. The crystals which are of the cubic type are mounted on a turntable which can be rotated by means of a handle external to the chamber. Associated with the crystal circuit is a pentode isolator valve, whose output circuit can be adjusted to select the 3rd or 4th harmonic of the various crystal frequencies, the harmonic frequencies so selected coming within a frequency range of 134 kc/s to 100 kc/s from which the final frequencies will be selected, after multiplication, through the harmonic amplifier.

Harmonic Amplifier. In order to provide for a quick wave-change five independent frequency-multipliers and amplifiers are provided forming the harmonic amplifier unit, the connections of which are shown in Fig. 310. The first three stages of each unit employ pentode valves and are used for frequency multiplying from the selected crystal frequency, but the last three stages employ triode valves and are amplifying stages, arranged to deliver an output of about 120 watts on the final frequency. A selector switch enables any of the individual circuits to be connected through a feeder to the intermediate amplifier, shown in Fig. 311. Four of the harmonic amplifier circuits are set up for the four spot waves, whilst the fifth is available to work with the flexible-frequency master oscillator. The whole of this apparatus is built in duplicate.

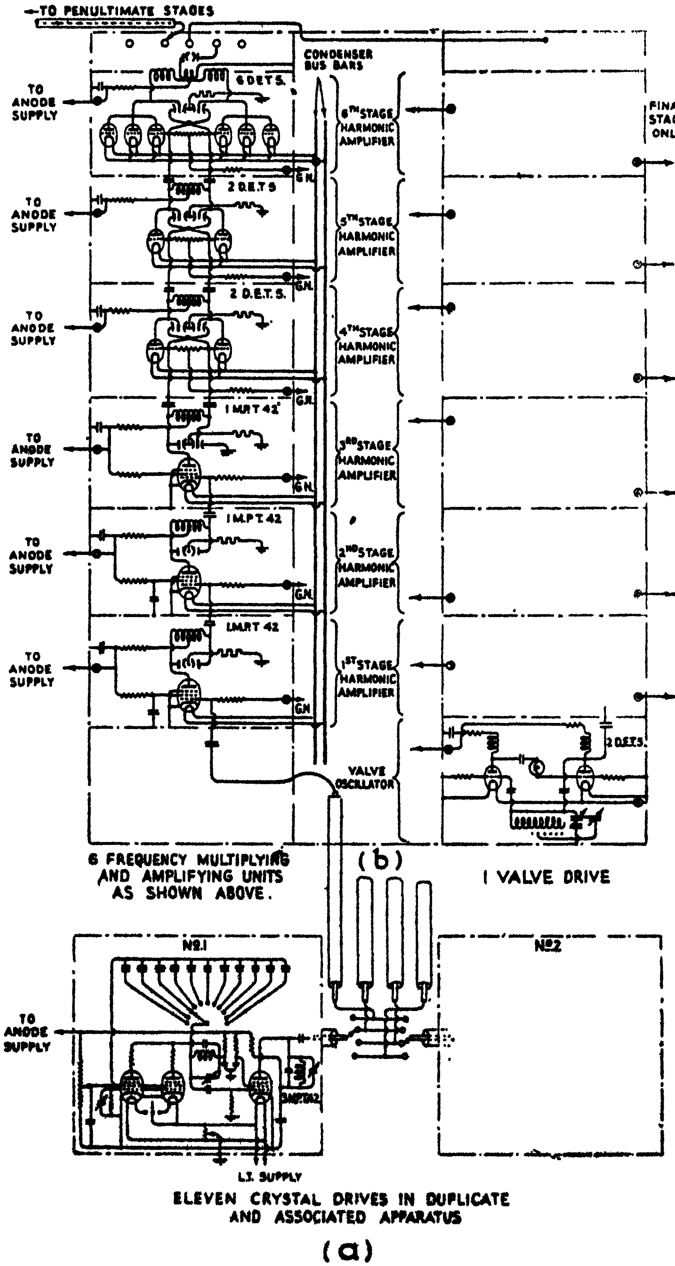

TO PENULTIMATE STAGES
CONDENSER BUS BARS
TO ANODE SUPPLY
6 D.E.T.S.
6TH STAGE HARMONIC AMPLIFIER
FINAL STAGE ONLY
2 D.E.T.S
5TH STAGE HARMONIC AMPLIFIER
2 D.E.T.S.
4TH STAGE HARMONIC AMPLIFIER
I M.P.T 42
3RD STAGE HARMONIC AMPLIFIER
2ND STAGE HARMONIC AMPLIFIER
1ST STAGE HARMONIC AMPLIFIER
G.N.
2 D.E.T.S.
VALVE OSCILLATOR
6 FREQUENCY MULTIPLYING AND AMPLIFYING UNITS AS SHOWN ABOVE.
(b)
I VALVE DRIVE
No.1
No.2
L.T. SUPPLY
ELEVEN CRYSTAL DRIVES IN DUPLICATE AND ASSOCIATED APPARATUS
(a)

FIGURE 310.

Intermediate and Final Amplifier. It would obviously be quite uneconomic to duplicate the high-power stages in the same manner and a normal switching system does not give a sufficiently compact lay-out where large powers are concerned, owing to the length of connecting leads involved. Previously adjusted "plug-in" coils could be used but they

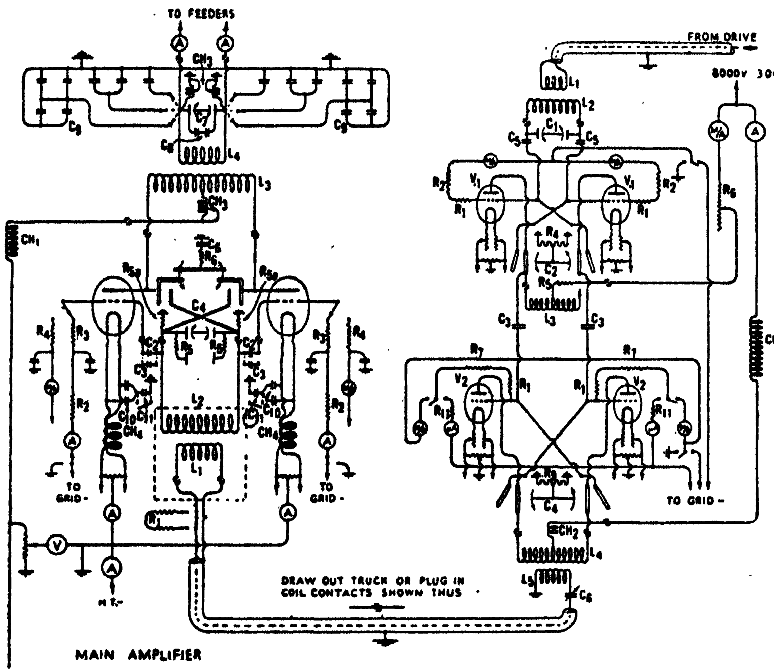


FIGURE 311.

are an inconvenient and inelegant solution, and the adoption of a turntable limits the number of circuits that can be pre-set. The solution in the transmitter being described is an ingenious one. A railway track is provided behind the transmitter and the anode and grid inductances are mounted on trucks together with the complete coupling circuit to the feeder. As many sets of circuits as are considered necessary can therefore be provided, and the method has the advantage that none of the idle circuits remain in the high frequency field of the transmitter, and the circuits not in use can be re-set to any new

frequency whilst the transmitter is working. When the appropriate truck is run up into the transmitter, the track ensures exact alignment of the contact system which is of a special design of a vice-grip type ensuring a very low-resistance contact. This arrangement of circuit trucks is employed for the last three stages of the transmitter, but whereas the final stage has a truck to itself, the two previous stages have circuits which are mounted one above the other and thus share a truck. A view of one of the trucks is shown in Fig. 313.

An unusual mechanical feature of design is that the movement of variable condensers, feeder coupling, etc., are controlled by a hydraulic ram, this giving a very smooth control and making the position of the various components quite independent of the control-lay-out of the front-panel.

The amplifier circuits (Fig. 311) are conventional triode balanced bridges as discussed on page 279. Thus considering the final stage (Fig. 311) L_1, L_2, tuned by its condenser, is the grid input stage; L_3, the output circuit inductance tuned by its condenser, the bridge being balanced by condensers C_4. The grid reactance condensers (the purpose of which was discussed on page 281) are shown as C_2 and the filament reactance chokes and condensers are indicated by CH_4, C_{10} and C_{11}.

As mentioned in a previous chapter, the stability and efficiency of a short wave transmitter depends, not so much upon the electrical circuit adopted, as upon the mechanical lay-out. Thus, the valves and circuits together must form a symmetrical group to earth, stray capacities must be reduced to a minimum, connecting leads cut to a negligible length, and coupling between the grid and anode circuits eliminated. The anode-circuit tuning inductance must be kept as high as possible and therefore the circuit tuning-condenser as small as possible and all conductor surfaces at high potential must be sufficiently rounded to avoid brushing or torch discharges. Insulators must, wherever possible, be kept outside the high-frequency field, and the dimensions of all conductors must be sufficient to carry the large currents that will be obtained on short waves at high powers. Thus with the present set, the main amplifier H.F. circuit at 13.5 metres will have a peak voltage of 20,000 when delivering full unmodulated output,

and an R.M.S. current of 240 amps., rising to 300 amps. on 100% modulation. The grid H.F. current on this wavelength is of the order of 120 amps.

The mechanical lay-out of the main amplifier is shown in Figs. 312 and 313 which illustrates a number of the points

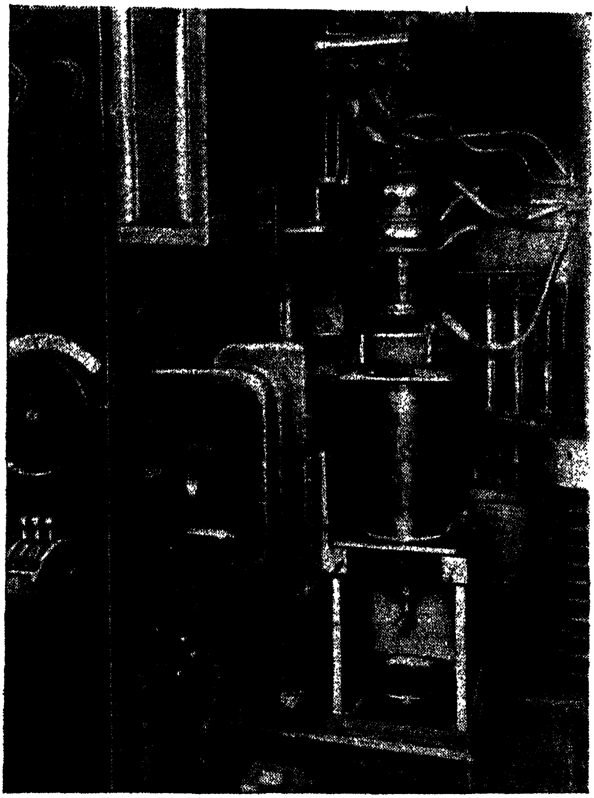


FIGURE 312.

mentioned. The circuit is built around two water-cooled triode valves of the CAT 17 type and from Fig. 312 it is observed that the valve jacket is bolted direct to a plate which serves both as one plate of the tuning condenser and as one plate of the balancing condenser, the other plates of those condensers being indicated in the figure. The thickness of these condenser plates is to be noted, the object being to make

possible a large radius of curvature at the edges, thus avoiding corona discharge. In spite of their apparent massiveness the actual capacities as measured across the diagonals of the bridge are quite small, namely 120 $\mu\mu F$, this representing the total capacity and thus requiring 4 μHys to tune the amplifier to the shortest wavelength, namely 13 metres.

FIGURE 313.

The grid tuned-circuit is behind the screening box shown in Fig. 312, which is cut away only to allow the grid input circuit to make connection to the grid seal of the valve, which takes the form of a complete copper ring having an ample current-carrying capacity for the large H.F. current passing to the valve grid. The valve and condenser assembly is mounted on a box-shaped insulating pedestal of mycalex.

The design of the grid and anode truck-mounted circuits

is seen from Fig. 313 which shows the 16-metre truck. The anode inductance consists of a group of horizontal tubes (four only in this case) cross-connected by tubes which can be slid to the position required for tuning, supported from the vertical busbars by mycalex insulation where necessary, the ends of the inductance being connected directly to the busbars, these latter being mounted on pedestals of mycalex. Between the turns of the anode inductance will be seen the output coupling inductance bolted to the tuning condenser which is supported on a separate pedestal. Examination of the layout will indicate what a small amount of insulating material lies within the H.F. field.

The grid-circuit tuning inductance is carried on two pairs of curved mycalex supports above the anode-circuit busbars, and in Fig. 312 will be seen the short length of cross-over arms from the balancing condensers rising vertically through central vertical column.

The particular transmitter built for the B.B.C. employs a series modulator, but the transmitter can be arranged equally well for Class B push-pull modulation. Both systems have been described in Chapter XIII, but whereas series modulation employing the cathode-follower circuit is almost free from distortion as it is a 100% negative feed-back circuit, it is usual to incorporate a negative feed-back circuit when a Class B modulator is used.

In the case of a series-modulated circuit, since the modulators operate at high power and in opposite phase to the H.F. amplifiers, the loading on the main power supply is constant, and in consequence there will not be any tendency to scintillate during modulation. But with a Class B modulator rapid fluctuations of load will be experienced at modulation frequencies and in consequence scintillation will occur unless precautions are taken to avoid it. These involve the use of filters in the anode circuits of the modulators to smooth out the A.C. components of the speech, and the use of low reactance supply circuits.

CHAPTER XVIII

HIGH-FREQUENCY THERAPEUTIC APPARATUS

For the benefit of readers unversed in medical terms, we may say that "therapy" is defined, in a delightfully vague manner, "as given to that branch of medicine which deals with the means to prevent or cure disease, or to control and lessen its evil results, if a cure is impossible." Thus, the homely hot-water bottle is probably the simplest piece of therapeutic apparatus, since it is well known that warmth is of the greatest assistance in practically every kind of ailment, and therefore any treatment which induces heat may claim to be called therapeutic in action.

Artificial methods for inducing heat within the body have been sought by the physician throughout the ages, but it was not until well after the beginning of the present century that the use of currents of radio frequencies became an established method. Of course for many years previously the effects of "Faradic" currents, so called by the medical profession, had been widely studied, and the "shock-coil" treatment for stimulating the nervous and muscular reactions had been in common use.

It was D'Arsonval who showed that the application of current having frequencies above 500 kc/s had no "Faradic" effect, but that such currents had a certain therapeutic action, although at the time he did not associate this therapeutic action with heat.

The recognition of the heating effect of such frequencies and the possibilities of their use appears to have been established first by Nagleschmidt, and he is responsible for introducing the now current term for this kind of heat treatment—diathermy—the name implying a "heating through" of the tissues, as distinct from a mere surface heating such as is produced by an external application of heat.

Until fairly recent times the only diathermy apparatus available has been designed to operate on wavelengths of the order of 200 metres (1·5 Mc/s). Recently, however, much shorter waves (down to about 2·5 metres 120 Mc/s) have been employed. Hence modern diathermy can be classified as:

(*a*) Medium-wave diathermy, also called conventional diathermy, or simply "diathermy."

(*b*) Short-wave diathermy (the name now generally considered to be most acceptable), which has also been described as "ultra-short wave therapy," "radiothermy," "parathermy," "neothermy," and "short diathermy."

It should be noted that the classification by wavelength (rather than frequency) is rather illogical because such wavelengths only exist in an unwanted by-product—radiation into space—but the terminology is so well established that it seems necessary to follow it.

It is not within the compass of this chapter to discuss the various medical applications of diathermy apparatus, which are treated in various books on the subject [1], [2], but it may be presumed that, in general, the heating produced is merely an aid to the natural process of the healthy tissue to rid itself of infection. In some cases where the lethal temperature of the infection is low, the heat produced has a most potent effect, but even when the heat cannot be raised to that temperature necessary to "kill," its action together with Nature's own reactions can effect a cure. And in the many cases where no cure can be obtained, such as those of rheumatic and muscular type, relief from pain can be effected in many cases.

The ordinary diathermy apparatus, in addition to being of use for straightforward heating treatment, is employed, to a less degree, for surgical operations, that is, for the excision of tissue, either by coagulation, desiccation (that is dehydrating the tissue), or by the use of an arc from the end of a pointed electrode which takes the place of the surgeon's knife. These uses, however, are more or less surface effects, and we shall not be concerned with them in the future discussion.

In 1925-26 short and ultra-short waves began to attract attention and since then a considerable amount of electro-

medical work has been done on wavelengths between 3 and 30 metres, mostly in Germany and America, and the literature on the subject is very extensive. Unfortunately, however, much of the earlier work of responsible electro-medical research workers was vitiated by the too hasty conclusions of men who either allowed sensational journalism to exploit their early efforts, or published the results of immature work with misleading conclusions. In spite of the similar experiences of early workers in medium-wave therapy and the well established fact that such currents produce only a heating effect, efforts appear to have been made to try to believe that ultra-short waves produce all kinds of specific effects other than heat, including atomic and molecular changes. It is curious how history has a habit of repeating itself. Dr. Cumberbatch observes in his treatise on diathermy, "The hypothesis that they (i.e. medium wave effects) are due, partly or wholly, to some other action of the current, as for example, 'massage of the protoplasm' by the oscillating ions, affords no guide to the selection of the maladies suitable for treatment. It must be remembered that the inability to find explanations—other than those which themselves require explaining—of the method of action of high-frequency currents on the body in the days when they were first used caused high-frequency treatment to fall into a 'slough of despond' from which it was only rescued when it was established that the therapeutic effects of these currents were due to the production of heat." It would almost appear that short wave therapy may be going through a similar "slough of despond."

Since the various specific effects attributed to ultra-short waves are now considered to be non-existent, there is no point in cataloguing them, but it may be of interest to give a few examples of literature, from a Continental source, which revolved around this research work a few years ago, and which can be offered without comment:

"The equilibrium of the wave radiation renders possible the building up of protoplasms. Protoplasm is nothing but simple dynamic machinery which is as little mysterious as a modern internal-combustion engine. If we wish to increase the activity of the brain, we must vary the wavelengths of the life rays in such a way that the percentage of the short

wave emanation increases with the oxidation process of the brain."

"The medical man of the coming century will be the radio mechanic of the human body, by this means he will be able to determine earlier the onset of a disease and be in a position to prevent it. On this basis it will be possible to effect an absolute cure of pulmonary tuberculosis and diabetes and completely to investigate the secrets of cancer."

"The complete disclosure of the secrets of cancer is said to be in progress in Berlin. It has now been found that ultra-short wave wireless waves exercise an annihilating effect on human cancer tissue."

"In the course of events therapeutics has arrived at short wireless waves. On the basis of previous indications the observations made probably signify a landmark in the domain of therapeutics. As the human body shows electrical oscillations, all those disorderly oscillations which proceed from a disease will be eliminated by the corresponding number of oscillations."

"*The Mysterious Power of the Wavelength of* 3·65 *Metres.* The rays which have penetrated the organism will according to their wavelength besides the thermic effect also exercise an annihilating effect on the cells of the ulcer; and this effect concerns us here. The effect is strictly dependent upon the wavelength, it turns upon minimum differences. Thus the wavelength of 3·65 metres is the most effective on cancerous cells, from 3·3 metres downwards, and upwards from 3·7 metres the wavelengths are without effect. Other diseases of course demand other wavelengths, and the different types of bacilli are also annihilated by different wavelengths. It is clear that the true value of ultra-short wave therapy will only develop when we know, when and with what wavelength work is to be done."

Unfortunately, in addition to journalistic efforts of the type quoted, which usually do no harm to a scientific cause, technical writings also appeared (and still do) which made claims to having produced results of the specific action of ultra-short waves which were not attributable to heat.

In this country such work on short and ultra-short waves had not made much progress, although what little had been

done gave no support to there being any action which could not be explained by heat, but in 1935 the British Empire Cancer Campaign considered the matter of such importance that they organised a major research into the various claims made, as it was clear that if confirmation of these results could be obtained a new and important physical aid to the general problem of disease was at hand.

The results of this research work, carried out by Dr. Frank Dickens[3, 5] and others, gave, however, a complete refutation of any of the claims for specific action, other than heating. These investigators pointed out that it should be possible for the discoverers of these effects to describe at least one experiment which could be repeated, and remarked that the inconclusive evidence which had been presented is a very insecure foundation for the rapidly growing belief in specific short wave therapy. It is of interest to record that much of the inconclusive evidence has since been withdrawn.

In a report to the Council of Physical Therapy (America), Dr. Franck Krusen[6] has given a concise summary of the position, which indicates very clearly that as far as is known at present short and ultra-short waves are a simple diathermic agent, that there is no bacteriological, physiological, or specific effects, and that it is doubtful if there is any selective action of the different frequencies even as regards heating effect, though there will be different problems connected with technical design and application.

This is not to say, however, that short-wave therapy is not of value. It most definitely is, but as far as present knowledge extends it must be regarded as a simple diathermic aid, and it thus comes to a clear-cut issue as to whether medium or short waves are better for such purpose.

This is a matter entirely for the physician to decide, although it appears to be generally agreed that the short wave diathermy is rather easier of application as no actual contacts have to be made to the patient, its heating effects are more uniform, they can be localised better and deeper penetration can be obtained. The chief difficulty appears to be in measuring the dosage.

We propose now to discuss the general problem from the diathermy point of view alone. If we agree that the application of heat to an affected area is the object of the treatment,

then the most important features of any apparatus will be the means for controlling the amount of heat induced, its direction to the affected part and, if possible, means for measuring the dosage.

The Human Body as a H.F. Circuit. The human body is, evidently, very inhomogeneous, and the capacity and resistance between different points on its surface will be widely different. The effective dielectric constant is, however, largely that of the saline solution which permeates the tissues and is approximately constant at about 80.

The effective resistance is much more variable and, moreover, is very difficult to measure in the case of living tissue. Measurements made on dead tissues will not be correct owing to the large influence that the permeating fluid will have in life. It is known, however, that fat has the highest resistance, whilst bone is fairly high. Skin, muscle, and the fleshy organs have comparatively low resistances, the lowest values occurring where the blood stream is heaviest. All resistances become somewhat less as the frequency is raised.

From the electrical point of view, the body is equivalent to a complex network of imperfect capacitances and resistances and its impedance will therefore be very different in a medium-wave system to what it is in short wave system. This is so because with medium waves it is the resistance of the body rather than the capacitance which determines the potential distribution and current flowing, whereas with short waves it is the capacitance of the body rather than the resistance which will determine the current flow.

Medium Wave Diathermy. In this case it is necessary for the electrode system to be in actual contact with the body. The series reactance of air-gaps between the body and the electrodes would be too high to allow of sufficient current flowing.

Although the average resistance of the body will determine the effective load-current flowing, the paths of the currents from one electrode to the other will depend upon the relative arrangement of the organs, i.e. whether they are virtually in series or in parallel. If they are in series, then the largest voltage drop will be across that part having the highest h.f. resistance and this part will in consequence become most heated. On the other hand, if they are in parallel, the part

having the least resistance will have the greatest heat produced in it, since in this case the current value through the low resistance path will be the greater.

Since the actual body is a complex arrangement of series and parallel circuits, one portion of the tissue between the electrodes often becomes hotter than the remainder, that is, selective heating takes place. Any alarming rise in temperature of one spot is largely prevented, however, by the action of the blood stream, which carries heat away and ensures its more even distribution.

Since the current will follow the low resistance paths wherever possible, it is found in many cases that heat cannot be induced

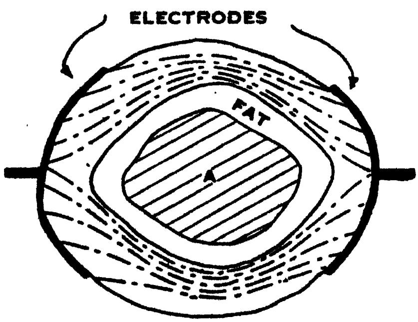


FIGURE 314.

in certain organs which are surrounded by a low resistance path. For instance, consider the example shown in Fig. 314, where A is an organ surrounded by a high-resistance fat layer, which is a non-conductor of heat, and surrounded again by a low-resistance, blood-stream area. It is clear that the h.f. current will be by-passed around the organ A, even though the organ itself is of a low-resistance nature.

Even if the body itself were homogeneous, this distribution of potentials and currents (and therefore of heating) would depend upon the shape and position of the electrodes one to the other. Thus, if the electrodes were inclined at an angle to one another, the greatest heating would occur where the electrodes were nearest. If one large electrode and one small one were used, heating would be greatest in the neighbourhood of the smaller one. This is true whether resistance or capacity is controlling the distribution.

Thus in surgical work a small area electrode is employed to produce such localised heat immediately beneath it that a volume of tissue immediately beneath the electrode can be raised to sufficient temperature to dehydrate the area, acting in a similar way to surgical cautery.

Thus the art of diathermy is largely that of applying the correct electrode system in the correct manner to achieve the best results for the particular part it is desired to heat, and to adjust the dosage for maximum therapeutic effect and not to overheat. As may well be imagined, this is not an easy matter and must be the result of long experience.

Medium-Wave Diathermy Apparatus. Since this is adequately treated in other text books and is hardly in place in this book, it will only be dealt with very briefly.

It is still quite common practice to employ a spark transmitter as the source of high-frequency currents. The considerations which have rendered the spark transmitter obsolete for wireless purposes do not necessarily rule it out for diathermy. It may, of course, produce very serious interference with communications, but the user has not—until recently at any rate—had to concern himself with this.

Spark apparatus will have the merit of simplicity and cheapness and less expense for replacements. It usually employs a multiple quenched-gap operating in coal gas, methylated-spirit vapour, or air.

Valve equipments are now coming into use, but in most cases apply an alternating voltage direct to the oscillator anode so that a 100% modulated wave is produced. A valve set is evidently simplified a good deal in this way, and as far as can be judged from the evidence available there appears to be no great disadvantage in the use of either the damped wave from a spark apparatus, or of a modulated wave for diathermy work.

It is certainly true that if the same "dosage" of a pure C.W. output and of a 100% modulated output is taken there are no appreciable differences of sensation or heating at frequencies of 1 Mc/s or higher, but it has been stated that, when the limiting frequency of 500 kc/s is approached, differences are perceptible, particularly when electrodes are in imperfect contact. There would appear to be more difference between

spark and modulated C.W. than between modulated C.W. and pure C.W.

That there is no apparent objection to the use of a modulated wave is presumably due to the fact that at the powers necessary to induce heat no part of the body acts as a rectifier. With powers sufficient to overload the tissue it is reasonable to assume that a rectifying action would be set up and low-

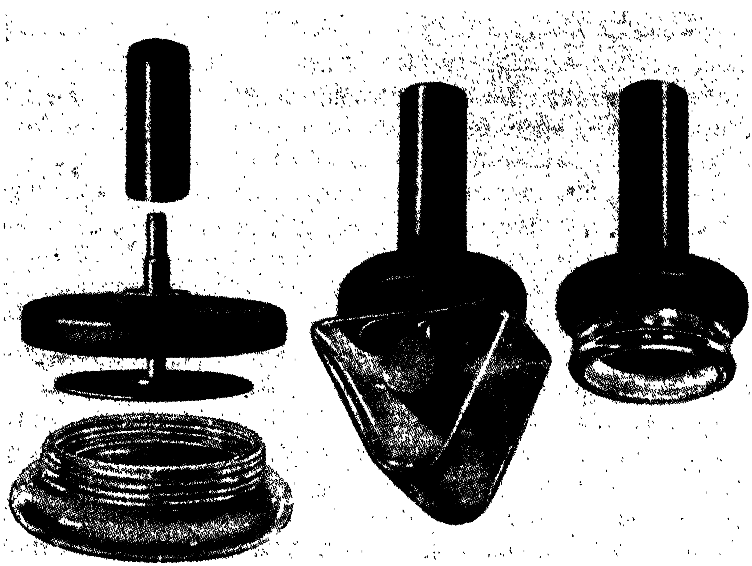


FIGURE 315.

frequency currents would then be present. In consequence, some physiological action would occur which might or might not be harmful.

Short-Wave Diathermy. We have already seen that in this case it is the capacity of the body which is of importance in settling the current and potential distribution. The capacity reactance between the electrodes is quite low at these high frequencies. The heating is produced, of course, because the tissues of the body have dielectric losses and also because they have finite resistance.

The reactance of air-gaps between the electrodes and the body will now be quite low, and because of this it is no longer necessary, or even desirable, to connect the electrode system of a short-wave diathermy apparatus directly to the patient, but it may be coupled through the medium of condensers of small value. Various forms of such condenser electrodes, designed by Dr. Schliephake, one of the earliest workers with ultra-short wave therapy apparatus, are shown in Fig. 315, and these are typical of electrodes now in general use. As can be seen from the figure, the condenser plates themselves are movable within the glass insulating-cap, so as to allow of the distance to the body being varied at will.

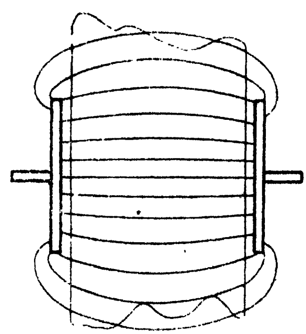


FIGURE 316.

We will now consider the patient circuit further. If we assume that the coupling of the electrode system to the patient is tight, as shown in Fig. 316, there will be no appreciable radiation field from the electrodes, and since the dielectric constant of most human tissue is the same ($\kappa = 80$), we shall get a more or less uniform displacement current between the condenser plates and hence across the patient, although of course the heating internally of the different organs will still be a function of the losses within these capacitative organs.

It is because the current now is a function of capacitance of the body and not resistance that short-wave diathermy produces this more uniform heating, as it penetrates the deep-seated organs without difficulty. The current is not by-passed around certain organs because a low-resistance path encircles them, as sometimes happens at the longer wavelengths. As with the longer wavelengths, however, the shape, size and disposition of the electrodes will influence the path and intensity of the field. Dr. Schliephake has stated that parallel electrodes well spaced from the body give the most uniform field throughout the body, although of course a strong radiation field will be present with such wide spacing and in consequence the field density within the body will be weaker than with

more closely spaced electrodes. He also finds that with the electrodes very close to the body greater heating takes place on the body surface, and that if one electrode is placed close to the body and one further away, the greatest heat is induced in that part of the body nearest the close electrode. This latter effect is possibly due in some measure to the distortion of field caused by radiation.

Radiation from ultra-short wave sets is, in the author's opinion, more important than has been fully realised in the past. Although radiation from the oscillator can be avoided by careful screening, with the type of spaced electrodes now used it is not possible to avoid radiation from the patient circuit with its attendant difficulties as regards the measurement of dosage, because radiation is not a constant factor. Thus if the electrodes are closely spaced, because the patient is almost a short-circuit capacitance, radiation will be negligible and in consequence any meter placed in the patient circuit will give a current reading in proportion to that passing through the patient. But if the electrode spacing is opened up, a radiation field is established so that the current measured is no longer representative of that flowing through the patient, the radiation being inversely a function of the patient capacitance and directly a function of the electrode spacing.

As there does not appear to be any merit in using ultra-short waves, it is suggested that it might be more desirable to standardise on some wavelength between 10 and 20 metres—that is, short enough to have the advantages of condenser electrodes, but long enough to be able to devise a useful dosage meter. The adoption of a common wavelength would enable statistical evidence to furnish accurate, even if relative, dosage figures for different therapeutic applications. At present, with the variety of wavelengths in use, it is almost an impossible task to correlate the performance of one set with another without considerable experience with each type of apparatus.

The short wave system also allows of efficient heating by means of a coil field, the action being similar to that of the eddy-current furnace. Thus instead of a condenser output, a flexible cable is connected across the output terminals and wound in a convenient form around the part of the body to be heated. Such a coil output, of which an example is shown in

a later section of this chapter, is particularly useful for the application of heat to the limbs. Here again, though, the question of dosage measurement is just as difficult on account of the radiation field that is established.

Short-Wave Diathermy Apparatus. As short and ultra-short wave therapy apparatus is not yet standardised, we propose to describe briefly several equipments of the valve type. We might mention in passing, however, that even on short waves, spark apparatus is still quite common, but in view of the fact that it must be regarded as obsolescent we do not propose to include a description of any sets employing a spark transmitter.

Short-wave technique lends itself to simple methods of energy transfer through feeder lines, and it may be more convenient to separate the patient circuit from the main transmitter as shown in Fig. 317, which shows a short wave set with a screened feeder coupling transmitter output to patient circuit. Such feeders may be of the concentric tube, or screened twin type, although the latter will usually be found more convenient, and they may be coupled to the output circuit either magnetically or capacitatively. The particular circuit shown in Fig. 317 is that of an experimental model employing two water cooled valves and operating over a wave range of from 2 to 8 metres, giving an h.f. output from 1·8 kW at 2·5 metres up to 3·4 kW on wavelengths of 6 metres and above. As can be seen from the figure the circuit is a conventional self-oscillator, the valves forming the main tuning

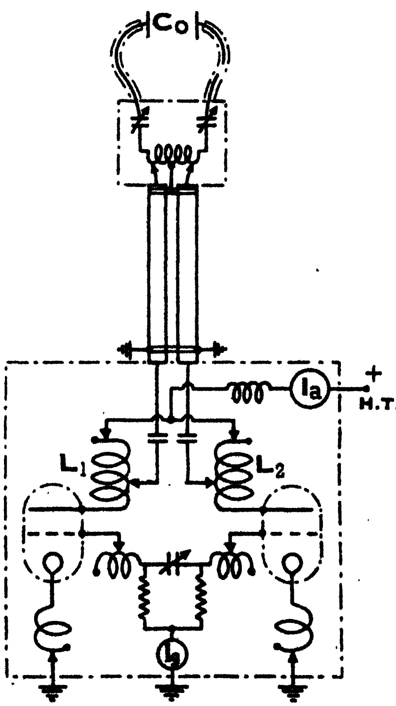


FIGURE 317.

circuit capacitance, L_1 L_2 the main tuning coil, and with the usual chokes in grid, anode, and filament circuits.

The "Ultratherm" Apparatus. A general view of this set, which is made by General Radiological Ltd., is shown in Fig. 318.

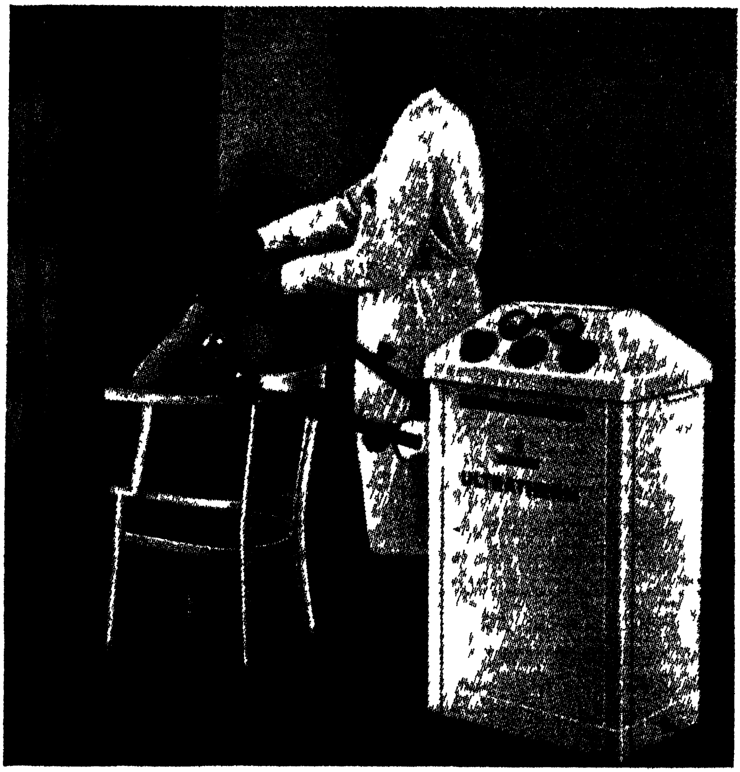


FIGURE 318.

It is designed to give an output of 300 watts at a fixed wave-length of 6 metres (50 megacycles), and Fig. 319 shows a diagram of connections of the apparatus. As may be seen from this, the set is a self-oscillating valve transmitter, fed from un-rectified A.C., the supply to the filament of the transmitting valve being also A.C., taken from a tertiary winding on the main power transformer. As was mentioned previously in

the main text, it does not appear to be very essential to use unmodulated h.f. current for diathermy work and hence it is a much more simple problem to build a power supply for a modulated output such as is shown in the diagram.

The tuned and coupled output circuit is shown at *LC*, across which are connected at 1, 2 the patient electrodes in parallel. These may be of the condenser-electrode type previously described, or a flexible-coil output may also be used. The condenser electrodes are shown in Fig. 318, held in position to a patient by a type of universal supporting arm.

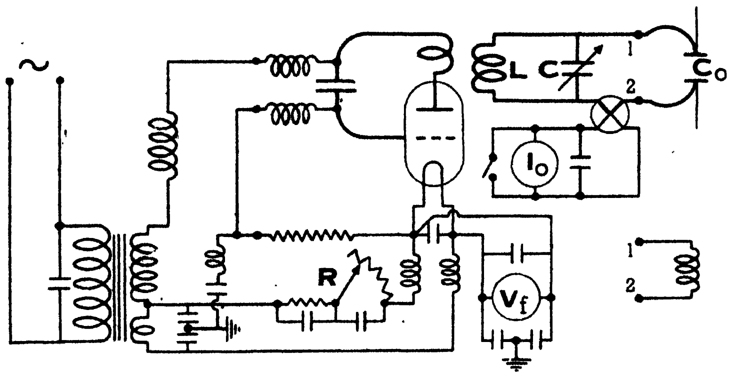


FIGURE 319.

A thermo-couple type of ammeter is included in series with the patient circuit but, as has been explained previously, such an instrument cannot give an accurate indication of dosage on ultra-short waves, on account of the radiation field present when the electrodes are altered with respect to the body. At the same time a skilled operator can, with experience, obtain a useful guide to dosage by means of such an instrument.

The control of power output is obtained, first by a careful setting of the patient electrode system, and secondly by a variation of the filament voltage to the oscillating valve. This is an unusual method of controlling output, and although most effective, as can be seen from Fig. 172, page 287, it is not a good method for the ordinary type of valve, as considerable secondary emission takes place from the grid. No doubt this point has been considered by the manufacturers, and possibly allowed for in the design of the valve.

Although the "Ultratherm" is nominally designed for diathermy treatment, the set can be used for the usual electro-surgical operations by the addition of appropriate fittings.

"Intertherm." In Fig. 320 is shown a general view of the "Major Intertherm" ultra-short wave diathermy set made by

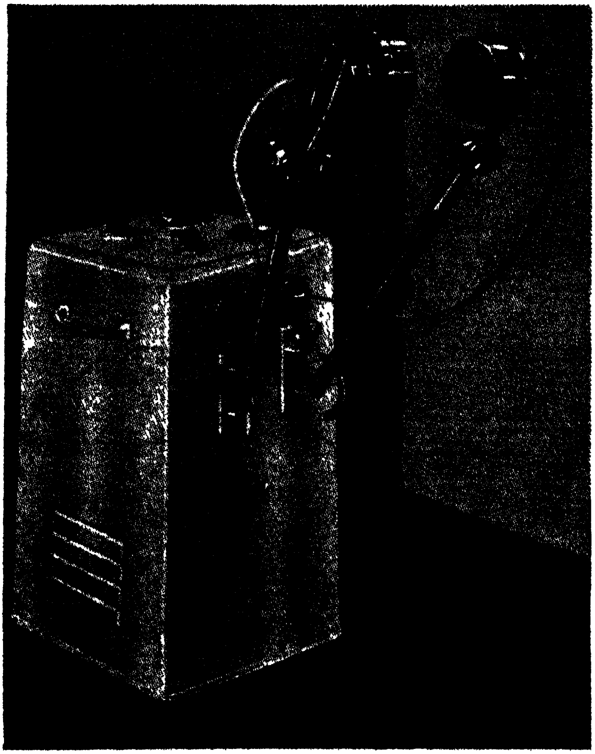


FIGURE 320.

Stanley Cox Ltd. This set operates on a wavelength of 6 metres (50 megacycles) and gives an output of 450 watts. Fig. 321 shows a simplified diagram of connections of the apparatus from which it is seen that the H.F. circuit is a straight-forward "push-pull" type of self-oscillator, fed from unrectified, unsmoothed, high voltage A.C., through a mains transformer T_1. The valve filaments are also heated from an A.C. supply obtained from a separate transformer shown at T_2.

The output circuit *LC*, coupled to the anode coil of the oscillator, has the patient terminals 1, 2 connected across it in parallel, and is suited either for condenser-electrode or inductive-coil application, as previously discussed. An H.F. meter is coupled into the patient circuit as a means of indicating when the patient has been tuned in correctly.

Power regulation is quite simply arranged for, by means of a tapped primary on the power transformer, the tap selection being by means of a radial switch *R*, and a thermal delay switch *S* included in the transformer primary circuit and controlled from a winding on the filament transformer ensures that no H.T. power can be switched on until the filaments are up to full voltage.

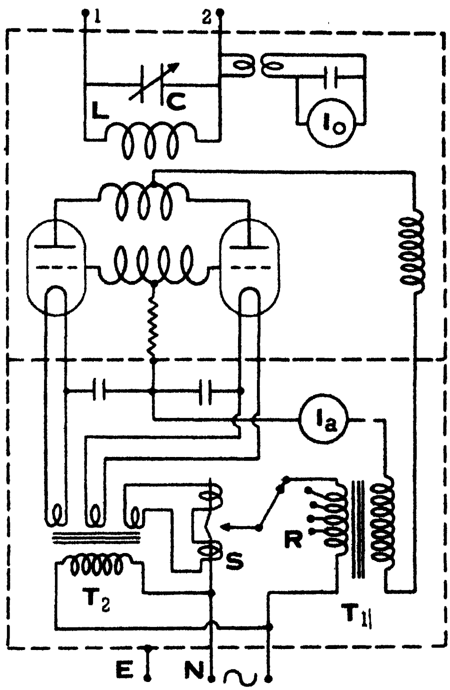


FIGURE 321.

The "Inductotherm." This is made by the Victor X-ray Corporation and is specifically designed for diathermy treatment, although an attachment can be provided for a full range of surgical work. The circuit for medical work is arranged for the application of high-frequencies by inductive coil coupling to the patient and no condenser output circuit is provided, the makers of the apparatus claiming that all forms of diathermic treatment can be adequately covered by their particularly adaptable form of coil.

The general view of the "Inductotherm" is shown in Fig. 322, which shows the special flexible coil. This coil, which is a multi-strand, flexible cable, heavily insulated, can be formed in a variety of ways to suit the particular treatment required.

Thus it may be folded into a closely-spaced helix, or into a pan-cake form, for intense local treatment; or into an open form, for a general distribution of heating effect, such, for instance, as may be required for inducing artificial fever.

A general diagram of connections of the apparatus is shown in Fig. 323, from which it is seen that the H.F. circuit is a self-oscillating transmitter, fed from an H.T. supply derived from unrectified A.C., the valve filament being supplied from a tertiary winding (not shown) on the power transformer.

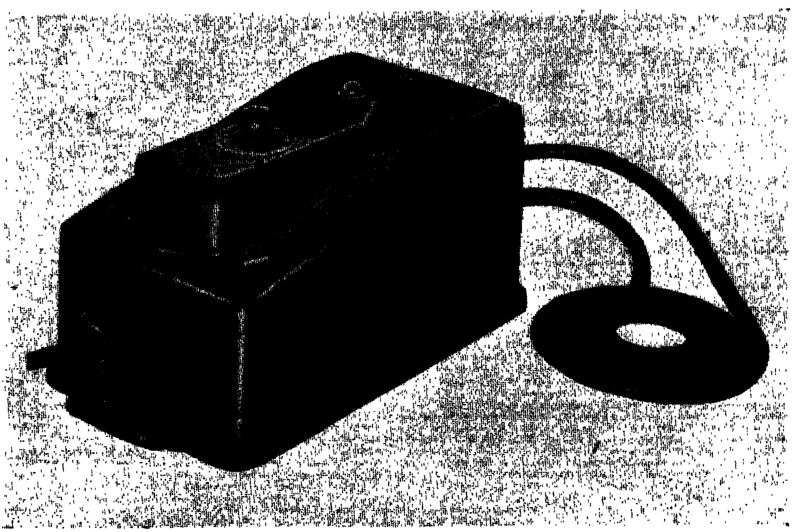


FIGURE 322.

The self-oscillator is a fairly conventional type of Colpitts circuit, the patient "Inductotherm" cable forming the output H.F. inductance. The average wavelength is of the order of 25 metres (12 Mc/s), the actual wavelength being determined by the particular shape in which the coil is formed and the relative position of the patient, but as has been explained in a previous section, within the limits previously specified, the actual frequency used for diathermy work is of little importance.

The method of controlling output is novel, and can be seen by considering Fig. 323. It is observed from this that the filament return lead is connected to a slider on a resistance R

in series with the secondary of the main transformer, although the D.C. grid lead is connected to the end of the secondary winding, point *B*. Thus as the slider is moved away from the point *B* the anode volts are decreased and the grid negative volts increased, and presumably a smooth reaction control is thus provided for the different power outputs.

No output meter is included, but a neon tube indicator gives a clear indication when the set is delivering power to the patient circuit. Experience has shown that such a meter is no measure of heat generated in the tissues and its inclusion is therefore liable to lead to misinterpretation and perhaps faulty treatment.

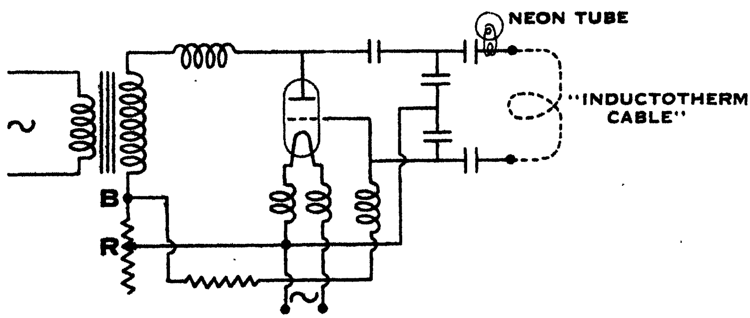


FIGURE 323.

Ultra-Short Wave Magnetion Set. The ultra-short wave set shown in Fig. 324 known as type UFG1, made by the Marconi Company, was designed by one of the authors for pathological research, and contains some interesting features. The set uses a split-anode, magnetron valve, made by the M.O.V., Hammersmith, and it is arranged to operate as a dynatron oscillator. Oscillating in this way (see page 378) it gives a high efficiency over its working wave-range, from 2·6 to 8·5 metres. The H.F. output obtained is 160 watts over the optimum part of the wave-range, that is from 3·5 to 5 metres, and somewhat less on other wavelengths.

The power supply unit consisting of a transformer *T*, Fig. 325, supplying power to two condensers C_1 and C_2 set for voltage doubling through two rectifiers R_1, and R_2 which are of the copper-oxide type. The valve filament is heated from A.C. and a separate transformer supplies current to the magnet

field windings *F* through a large capacity copper oxide rectifier in the secondary winding.

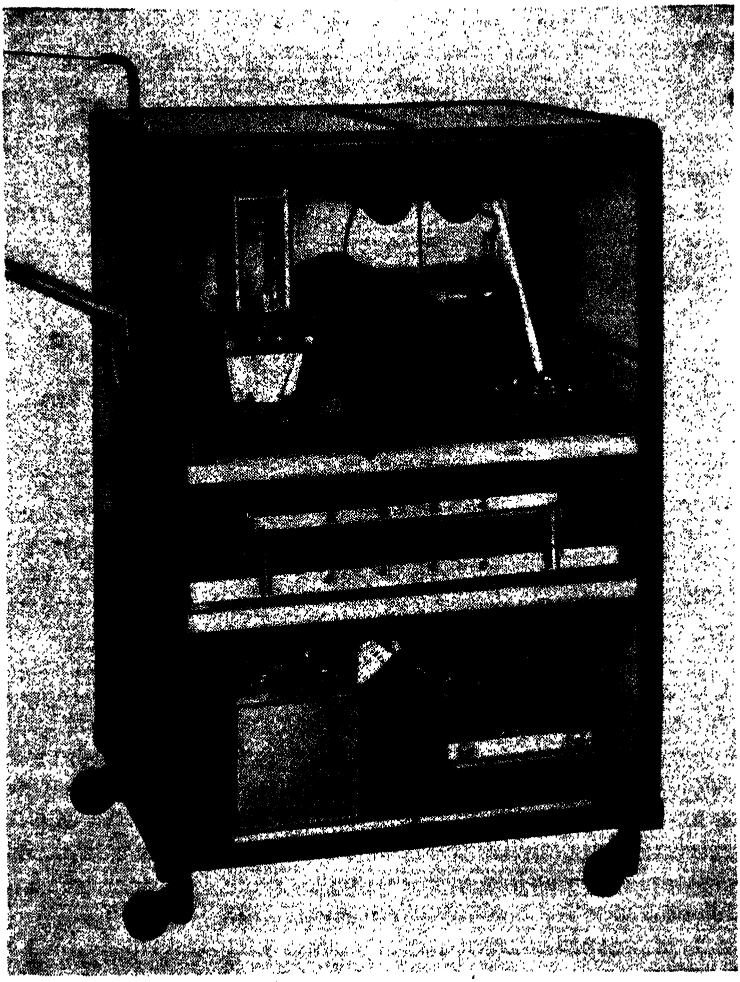


FIGURE 324.

The complete magnetron H.F. unit can be removed in order that it may be operated on a bench if desired. The magnet core is made of a "Lohmore" iron which will carry high flux densities without saturating, and the valve holder is so designed

that a universal movement is possible in order that the valve can be correctly positioned and orientated within the pole pieces to obtain even heating. The H.F. circuit consists of a variable inductance formed of " U " tubes, in parallel with the valve capacitance and with a small, variable, trimming condenser. The coil assembly is mounted on an insulating bar, and the wavelength changed by sliding the " U " tubes up or down a series of vertical members, minimum inductance being obtained by inverting the "U" tubes on their slides. The screened-twin feeder to the load is capacitatively coupled to the output through two variable condensers.

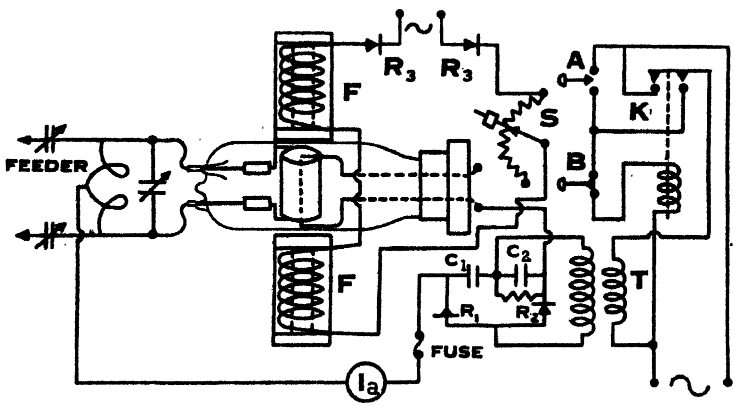


FIGURE 325.

A feature of the set is the neon-tube wavemeter which is incorporated in the panel of the H.F. unit. This wavemeter is removable by a simple half-turn and can be used either as an absorption meter without the neon-tube, or with the tube in place, calibrations both with and without the tube being given on the scales provided on the instrument.

For the safe operation of magnetron sets it is necessary to operate from a maximum field condition to one of lesser value and to ensure that this procedure is carried out automatically, a safety device is incorporated in the set consisting of a coil-operated contactor-unit *K*, the operating coil being controlled through contacts *A*, *B*, associated with the field-winding switch S. Thus until the field is brought up to maximum the

contactor remains open and is only made when the field is set to maximum. On reducing the field the contactor is held on by an auxiliary contact, but when reduced to " off " the main is tripped by a second pair of contacts, and the operation must be repeated.

SELECTED REFERENCES

(1) *Diathermy.* Elkin P. Cumberbatch. Heinemann.

(2) *Short-Wave Therapy.* E. Schliephake. The Actinic Press.

(3) *Annual Report, British Empire Cancer Campaign,* 1936. Page 163.

(4) *Short Wave Therapy.* Cancer Campaign Investigation. *British Journal of Physical Medicine.* R. G. Canti.

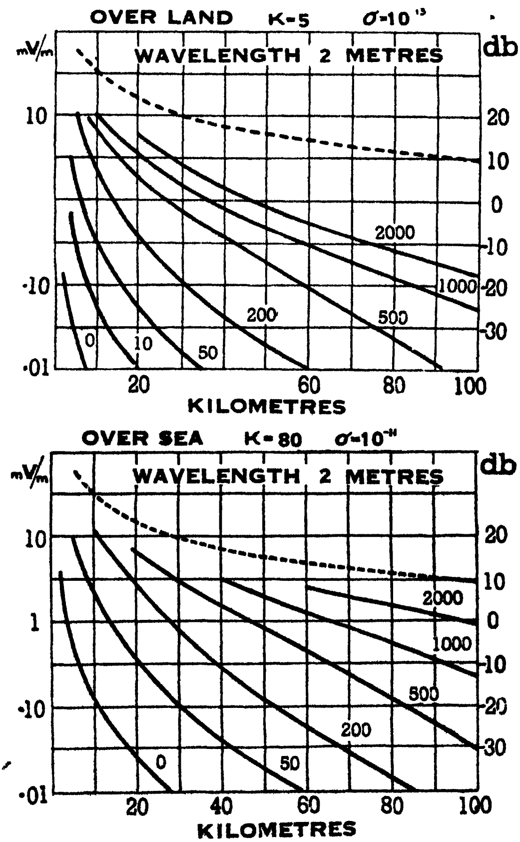
(5) *The " Specific " Action of Ultra-Short Wireless Waves.* W. E. Curtis, F. Dickens, and S. F. Evans. *Nature,* 138; July, 1936.

(6) *The Present State of Short-Wave Diathermy.* Frank H. Krusen. Report of the " Council on Physical Therapy." *The Journal of the American Medical Association,* April 16, 1938.

APPENDIX XI

PROPAGATION CURVES FOR WAVELENGTHS OF 2, 4 AND 6 METRES

FIGURES 326, 327 and 328 show curves from which the field strength due to a transmitter on the ground (using a wavelength of 2, 4 or 6 metres) at a receiver elevated at various heights can be obtained, both when transmission is over earth (of average σ and κ) or over sea. The heights (in metres) are indicated against each curve, values being given from 0 to 2,000 metres height. All curves are for 1 kW radiated from a half wave aerial, and for values σ and κ as indicated on the diagrams. Observe the greatly increased range obtained on all wave-



CURVES FOR 1 kW RADIATED.

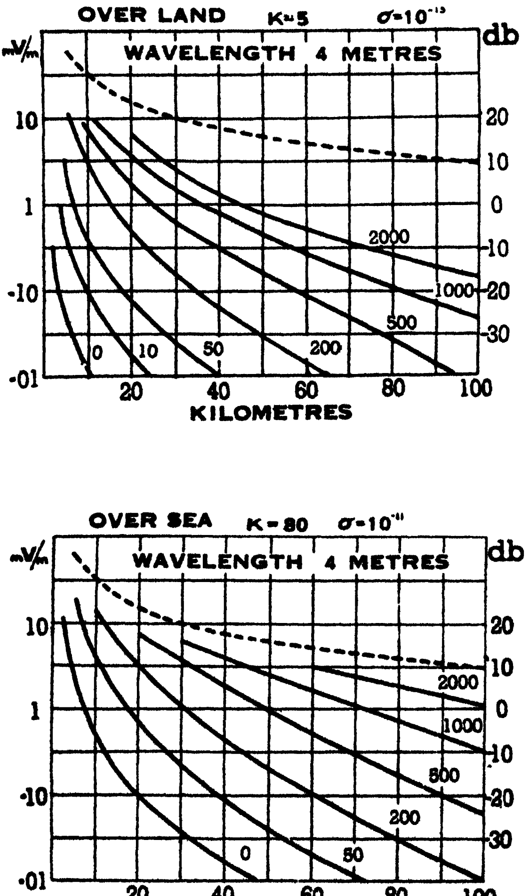
FIGURES BY CURVES INDICATE HEIGHT IN METRES OF TRANSMITTER (OR RECEIVER).

FIGURE 326.

lengths over sea, particularly when the aerials are both at zero height.

The set of curves given here is for shorter distances than the sample curves given in the main text, as it is considered that these distances are of more interest to the average experimentalist.

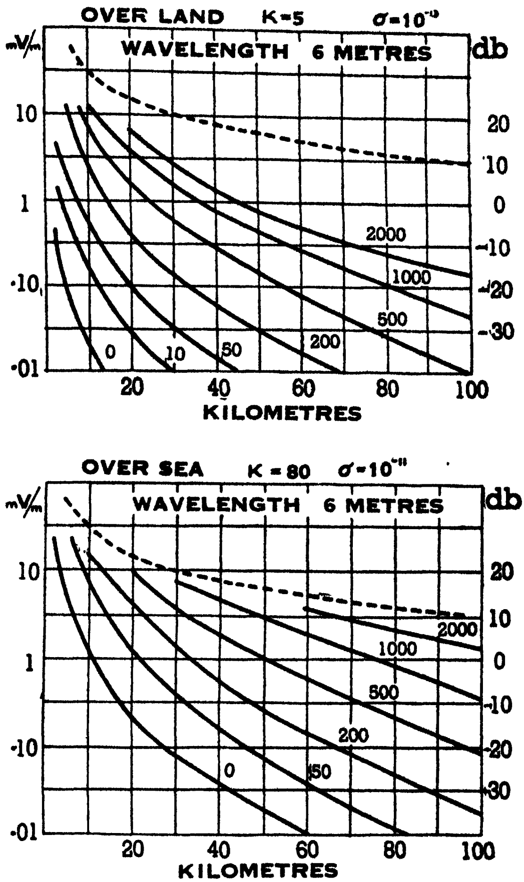
When both transmitter and receiver are elevated above earth, the gain in field strength can be obtained by using the supplementary curve of Fig. 329, or Fig. 40, page 70, in the manner explained on page 71. It should be noted, however, that when the height considered is below the critical values of h_e, the earth's constants have considerable effect and the appropriate curve given on Fig. 329 should be used. For great heights the curve on page 70 should be used. In both cases it is necessary to apply the correction figures given below the curves for sea and land.



CURVES FOR 1 kW RADIATED.

FIGURES BY CURVES INDICATE HEIGHT IN METRES OF TRANSMITTER (OR RECEIVER)

FIGURE 327.



Curves for 1 kW Radiated.

FIGURES BY CURVES ARE HEIGHT IN METRES OF TRANSMITTER (OR RECEIVER).

FIGURE 328.

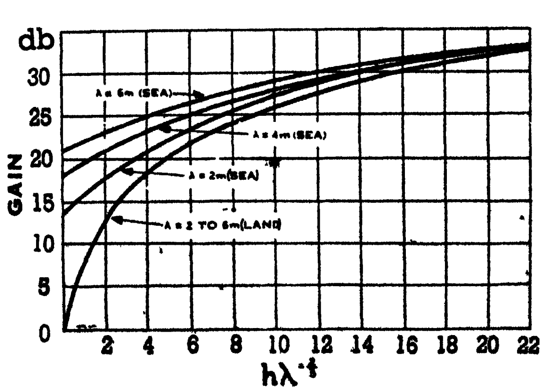


FIGURE 329.

The following are the figures in decibels to be subtracted:—

Wavelength (metres)	2	4	6
db to be subtracted			
Over Land	0	2·2	3·7
Over Sea	13·6	18·0	20·8

APPENDIX II

OPTIMUM WAVELENGTH FOR COMMUNICATION OVER DISTANCES BETWEEN 500 AND 5,000 KILOMETRES

THE following curves are suitable only for distances between the limits mentioned above, and give only an approximate result, since they average the conditions over any given route.

Daylight Conditions. When daylight spreads over the route between the stations being considered, both bending and

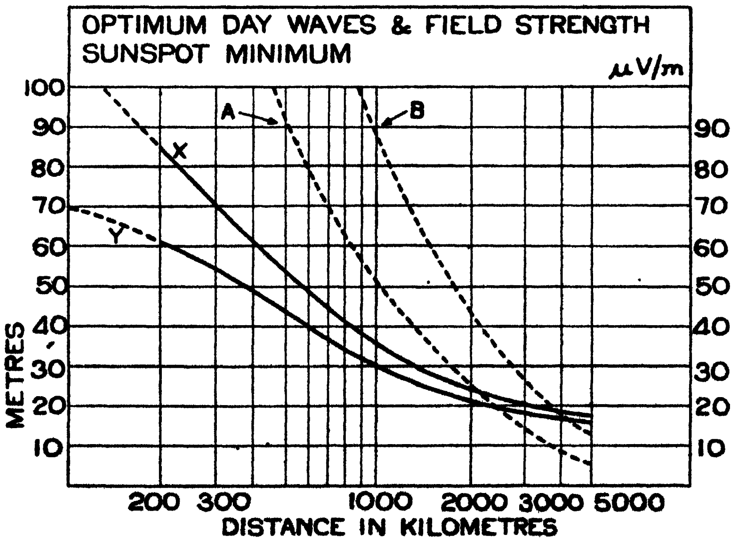


FIGURE 330.

attenuation have to be considered. Since the attenuation is least on the shorter wavelengths, the optimum wavelength will be one which is just above the skip wavelength for the distance considered. For any given distance, the day wavelength limits will be found between the curves *XY* Fig. 330. These curves apply to routes across temperate latitudes and

during the sunspot minimum years. For tropical routes at all times and over routes in temperate latitudes during sunspot maximum periods, the wavelength selected from Fig. 330 should be reduced by 20%.

Having obtained the optimum wavelength for the distance required, the field strength for 1 kW radiated can be found by reference to the curves *A* and *B*, where these letters have the same significance as in the main text, namely sunlight and weak sunlight.

Night Conditions. For communications over routes in darkness, the attenuation on all wavelengths is small, and the conditioning feature is bending, which is small. Fig. 331 shows a group of limiting wavelength contours (i.e. for the longest distance) for different seasons, latitudes, and local times after sunset, both latitude and time being considered at a point midway between the stations between which communication is desired. The midpoint time is found from the difference of longitude, namely 1° of longitude difference is equal to four minutes of time (remembering that longitude east, Greenwich time least), and the difference of latitude must be estimated knowing the latitudes of the stations concerned and their distance apart.

The time after sunset for the midpoint of the route can be found from Table I which correlates these local sunset times for the different seasons and latitudes.

The curves of Fig. 331 enable us to find the wavelength above which the bending is insufficient to return a ray to earth over the longest possible distance. Having found this limiting wavelength, or what is virtually the maximum electron density at the centre part of the route, reference can then be made to the skip curves of Fig. 55 page 99 to determine the minimum wavelength for the particular distance considered. This wavelength will of course be greater since the contour lines of Fig. 331 represent the limiting conditions for the most oblique angle of incidence.

Having selected a wavelength that satisfies the boundary conditions, the probable field strength can be ascertained from Table II, the field strength being approximately the same for all wavelengths between 25 and 50 metres. This is so, because it is the bending of the ray alone which determines the com-

TABLE I.

Local Sunset Times (24 hours) at various Latitudes.

Northern Hemisphere.	LATITUDE.							Southern Hemisphere.
	0°	10°	20°	30°	40°	50°	60°	
Jan. 1—Dec. 12	1800	1745	1730	1700	1630	1600	1445	June 10—July 3
Jan. 21—Nov. 22	1800	1745	1730	1715	1645	1615	1515	May 21—July 24
Feb. 8—Nov. 3	1800	1745	1730	1715	1700	1645	1600	May 1—Aug. 12
Feb. 17—Oct. 26	1800	1745	1745	1730	1715	1700	1630	April 22—Aug. 22
Feb. 28—Oct. 14	1800	1800	1745	1745	1730	1730	1700	April 11—Sept. 2
Mar. 13—Oct. 1	1800	1800	1800	1800	1745	1745	1730	April 1—Sept. 15
Mar. 21—Sept. 24	1800	1800	1800	1800	1800	1800	1800	Mar. 21—Sept. 24
April 1—Sept. 15	1800	1800	1800	1800	1815	1815	1830	Mar. 13—Oct. 1
April 11—Sept. 2	1800	1800	1815	1815	1830	1845	1900	Feb. 28—Oct. 14
April 22—Aug. 22	1800	1800	1815	1830	1845	1900	1930	Feb. 17—Oct. 26
May 1—Aug. 12	1800	1815	1830	1845	1900	1915	2000	Feb. 8—Nov. 3
May 21—July 24	1800	1815	1830	1845	1915	1945	2045	Jan. 21—Nov. 22
June 10—July 3	1800	1815	1830	1900	1930	2000	2115	Jan. 1—Dec. 12

nunication. The received signal will, in general, be fading and ;he field strengths given in Table II are the highest values to vhich the field strength is likely to rise. These field strengths ıre for 1 kW radiated from a half wave aerial, so that with ımaller or greater powers, the necessary correction must be nade for this.

TABLE II.

Approximate Night Field Strength values for 1 kW radiated (peak values) for all wavelengths between 25 and 50 metres from half wave aerial.

Distance. Kilometres.	Micro-Volts per metre.
500	1000
600	500
700	400
800	300
900	260
1000	225
1500	155
2000	108
2500	84
3000	67
3500	56
4000	47
4500	41
5000	36

Example. Select a suitable wavelength for communication ›etween London and Stockholm at 9 p.m. (2100) London, on 10th)ecember. Obtain the field strength for 100 watts radiated from half wave aerial.

The mid-latitude is approximately 56° and the distance 1,500 ms. The longitude difference between London and Stockholm is 8° and the midpoint time which is 9° E of London is 36 minutes ıter than 9 p.m., say 9.30 p.m. To find the hour this is after unset, refer to Table I. This gives us for a latitude of 56° for 2th December a time of 1500, and thus the mid-latitude time is ½ hours after sunset.

Referring now to Fig. 331, Winter, we see that the limiting 'avelength for latitude 55°, 6½ hours after sunset is 32 metres, and y reference to Fig. 55, for a distance of 1,500 kms, we follow the

32 metre contour, curve 4, which gives for this distance a minimum wavelength for communication of 50 metres.

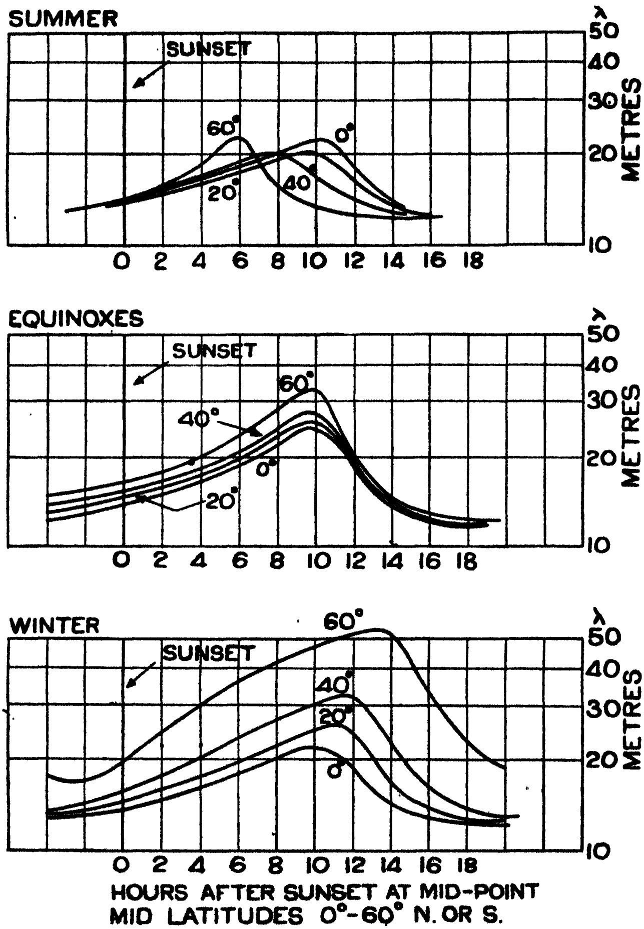


FIGURE 331.

From Table II, the peak field strength for 1 kW radiated is seen to be 155 micro-volts per metre, but since the power used is only 100 watts, the field strength will be:

$$\sqrt{\frac{100}{1000}} \times 155 = 45 \text{ microvolts per metre.}$$

APPENDIX III

CALCULATION OF THE CHARACTERISTIC RESISTANCE OF FEEDERS FROM DIMENSIONS

(1) **Twin Parallel Wires.** The capacity per cm. between two parallel wires far removed from other conductors is

$$\frac{\kappa}{4\log_e \frac{d}{r}} \text{ electro-static units}$$

where d=distance in cms. between centres of wires, r is radius of wires (cms.)

$$\text{or} \quad \frac{1}{4 \times 9 \times 10^{11} \times 2{\cdot}303 \log_{10} \frac{d}{r}} \text{ farads (in air), or}$$

$$\frac{10^{-10}}{828 \log_{10} \frac{d}{r}} \text{ farads.}$$

The inductance per cm. of two wires forming a loop is $4\mu \log_e \frac{d}{r}$ E.M. units

$$= 4 \times 1 \times 2{\cdot}303 \times 10^{-9} \log_{10} \frac{d}{r} \text{ henries (in air)}$$

$$= 9{\cdot}212 \times 10^{-9} \log_{10} \frac{d}{r}$$

$$\text{Hence } R_0 \text{ is} \quad \sqrt{\frac{9{\cdot}21 \times 10^{-9} \log_{10} \frac{d}{r}}{\frac{10^{-10}}{828 \log_{10} \frac{d}{r}}}} = 276 \log_{10} \frac{d}{r} \text{ ohms}$$

This formula is accurate down to a ratio d/r of 4/1, and Fig. 332 shows a curve of R_0 plotted against the ratio d/r, obtained by using the above formula, but corrected for ratios below 4/1.

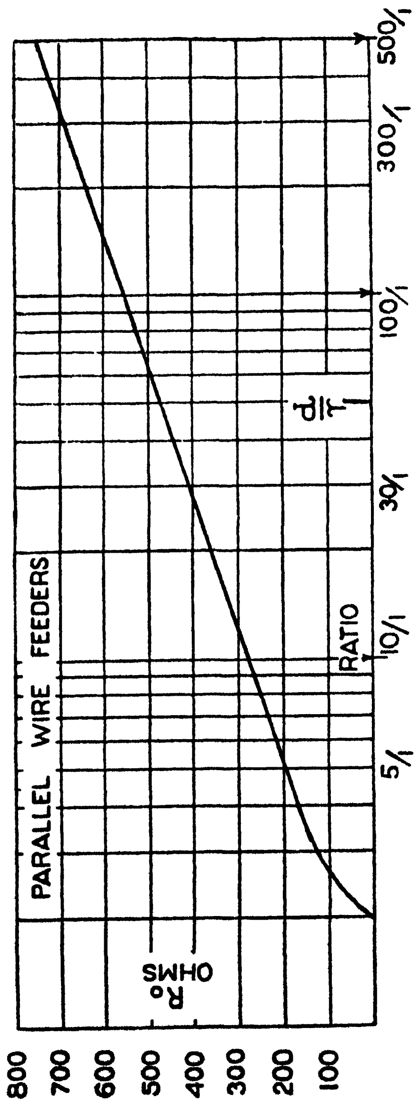
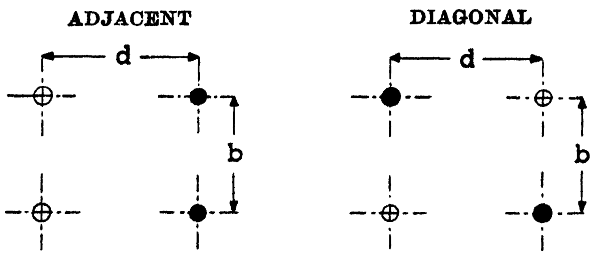

PARALLEL WIRE FEEDERS
R_o OHMS
800
700
600
500
400
300
200
100
5/1
10/1
RATIO
30/1
$\frac{d}{r}$
100/1
300/1
500/1

FIGURE 332

Four-Wire Feeders. With four-wire feeders, the wires may be run either with the go and return wires adjacent, as shown in Fig. 333a, or diagonally, as shown in Fig. 333b, where d is the horizontal distance between wires, b the vertical distance, and r the wire radius.

From a knowledge of inductance and capacity of the wires, the surge impedance can be derived for either type as with the twin wire feeder, and the values obtained are as shown by the formulæ below each figure.



(a) (b)

FIGURE 333.

ADJACENT

$$R_0 = 138 \log \frac{d\sqrt{d^2+b^2}}{b\,r}$$

When $d = b$

$$R_0 = 138 \log \sqrt{2}\,\frac{d}{r}$$

DIAGONAL

$$R_0 = 138 \log \frac{d\,b}{r\sqrt{d^2+b^2}}$$

When $d = b$

$$R_0 = 138 \log \frac{d}{r\sqrt{2}}$$

The curves for four-wire feeders are shown in Fig. 334, from which it is observed that both have a smaller surge impedance than the twin wire; that the values for R_0 approach equality when the ratio d/b is very small; that the curves for the adjacent spacing rise to become asymptotic to values for a twin wire when the spacing of d/b is large; and that an increase of the ratio d/b decreases the surge impedance of the diagonal spaced feeder.

In general this will mean that for diagonal spaced feeders it will be desirable to keep up the ratio d/b, whereas for adjacent spacing it should be kept down, but in both cases, a value of d/b of unity is a good starting point for consideration.

Power Ratings of Open-Wire-Feeders. The choice of feeder dimensions is governed by both practical and technical considerations. It is clear that in a transmission line the twin wires, when stretched between supports, must be far enough apart to avoid the risk of swinging together in high winds: about 6 in. separation is the safe minimum, for low powers, while 10 in. or 12 in. is the normal limit in the other

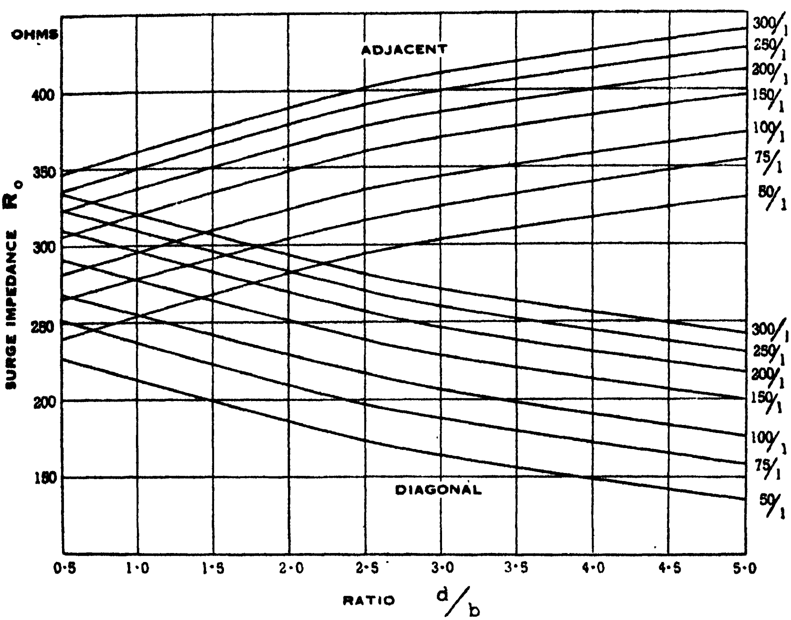


FIGURE 334.

direction, after which the cross arms, etc., become rather clumsy. In the case of reception the conductors require to be brought as close together as possible, in order to reduce the pick-up of other and unwanted signals; normally about 1¼ in. is a satisfactory separation, and as this is so small it is necessary to introduce separators every ten feet, more or less, in order to keep the wires apart. Returning to transmission lines, the main technical consideration is the need to keep voltages down on high powers, in order to reduce the tendency to torch discharge and flash-overs; this can only be effected by making the impedances low, hence the use of four-wire lines

for very high powers. A more direct means of overcoming the effects of high voltage is to increase the diameter of the wire as far as economical considerations will permit, since, very roughly speaking, the tendency to torch discharge and flash-over varies inversely as the conductor diameter. Of course, increasing the diameter of the conductor reduces the surge impedance, but this effect is comparatively slight, and must not be confused with the main reason. Generally speaking No. 6 gauge wires separated 12 in. represents the upper practical limit for twin wire feeders, and this size of feeder will be good for 50 kW telephone carrier power, provided the feeder is properly terminated.

It is not economical to use four-wire feeders except for high powers, above 50 kW. For 50 kW we could use 10 gauge wire, and for 100 kW 6 gauge.

Table I gives an indication of the telephone power ratings of various wires for a twin feeder and the spacing of various gauge wires for various ratios of d/r.

TABLE I

Gauge of Wire	Radius r″	d″ for ratios d/r of: 300/1	250/1	200/1	150/1	100/1	75/1	50/1	Power Rating kW	Recommended Impedance
18	·024	7·2	6·0	4·8	3·6	2·4			*Receiver*	550
16	·032	15·6	8·0	6·4	4·8	3·2	2·4		·25	600
14	·040		10·0	8·0	6·0	4·0	3·0	2·0	1·0	,,
12	·052		13·0	10·4	7·8	5·2	3·9	2·6	5·0	,,
10	·064			12·8	9·5	6·4	4·8	3·2	25·0	,,
6	·096				14·5	9·6	7·2	4·8	50·0	570
0	·162					16·2	12·0	8·1	80·0	530
R_0 =		685	660	635	600	550	520	475		

Note.—d and r in *inches*.

Example.

1. Find a suitable feeder for 2 kW to match a 550 circuit.

From curve 1, a 550 ohm feeder gives a ratio of d/r of approximately 100/1. Thus a 12 gauge pair spaced 5·2 inches would be suitable.

(2) **Concentric Tubes.** Capacity per cm. $= \dfrac{1}{2 \log_e \dfrac{r_2}{r_1}}$. E.S. units,

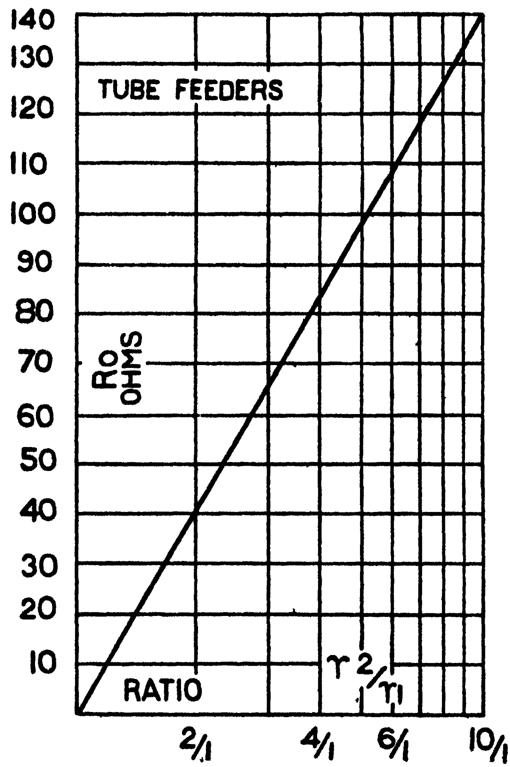


FIGURE 335.

where r_2 is inner radius of outer tube, and r_1 is outer radius of inner tube, and dielectric is air.

$$\text{Capacity per cm.} = \frac{10^{-11}}{9 \times 4{\cdot}605 \log_{10} \dfrac{r_2}{r_1}} \text{ farads}$$

$$= \frac{10^{-11}}{41{\cdot}45 \log_{10} \dfrac{r_2}{r_1}} \text{ farads}$$

Inductance per cm.

$$= 2 \log_e \frac{r_2}{r_1} \text{ E.M. units}$$

$$= 2 \times 2{\cdot}303 \times 10^{-9} \log_{10} \frac{r_2}{r_1} \text{ henries}$$

$$= 4{\cdot}606 \times 10^{-9} \log_{10} \frac{r_2}{r_1} \text{ henries}$$

$$\therefore R_o = \sqrt{\frac{4{\cdot}606 \times 10^{-9} \times 41{\cdot}45}{10^{11}}} \log_{10} \frac{r_2}{r_1} \text{ ohms}$$

$$= 138 \log_{10} \frac{r_2}{r_1}$$

Fig. 335 shows R_o plotted against the ratio $\frac{r_2}{r_1}$

Example. Find R_o for tubular feeder if outer dia. $= 3\frac{1}{2}''$, and inner dia. $\frac{7}{8}''$.

$$\frac{r_2}{r_1} = \frac{3{\cdot}5}{{\cdot}875} = 4{\cdot}1/1, \ R_o = 84 \text{ ohms.}$$

APPENDIX IV

FEEDER THEORY WHEN R AND G ARE ALLOWED FOR

THE complete theory for a line when conductor resistance R and leakage conductance G between conductors are allowed for is developed in many text-books and the results most often required in radio-frequency feeders are set forth here for reference.

The characteristic impedance becomes $\sqrt{\dfrac{R + j\omega L}{G + j\omega C}}$, whilst the propagation constant (the equivalent of m in our equations on page 131) is $P = \sqrt{(R + j\omega L)\ (G + j\omega C)}$. If P be split up into real and unreal parts a and β, then

$$a = \sqrt{\frac{1}{2}\left\{\sqrt{(R^2 + \omega^2L^2)\ (G^2 + \omega^2C^2)} - (\omega^2LC - RG)\right\}}$$

a is called the attenuation constant, because the maximum value of the voltage at any point x along an infinite (or correctly terminated) line is related to the maximum value of the generator voltage by the equation $E_x = E_G e^{-ax}$ . . (1)

and similarly $I_x = I_G e^{-ax}$ . . (2)

The loss in a correctly terminated line of length l is therefore given by:

$$\text{Loss in decibels} = 20 \log_{10} \frac{I_G}{I_R} \text{ or } 20 \log_{10} \frac{E_G}{E_R}$$

$$= 20 \log_{10} e^{al} = 20 \times 0{\cdot}4343\ al$$

$$= 8{\cdot}686\ al \quad . \quad . \quad . \quad . \quad (3)$$

If the frequency be so high that $\omega^2L^2 >> R^2$ and $\omega^2C^2 >> G^2$

then $a = \dfrac{R}{2}\sqrt{\dfrac{C}{L}} + \dfrac{G}{2}\sqrt{\dfrac{L}{C}}$

and Z_0 becomes $\sqrt{\dfrac{L}{C}}$, that is, the R_0 of our simplified analysis

of Chapter VI, so that $\alpha = \frac{R}{2} \cdot \frac{1}{R_0} + \frac{G}{2} R_0$ . . . (4)

or, if G is neglected, $\alpha = \frac{R}{2 R_0}$ . . . . (5)

Russell has derived a formula for the A.C. resistance of concentric tubes which for the high frequencies become

$$R_f = \sqrt{\mu f \rho}\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \text{ C.G.S. (E.M.) units per cm.} \quad (6)$$

where r_1 radius of inner conductor in cms.
,, r_2 ,, outer ,, ,,
If the conductor is copper (resistivity 1·7 microhms per cm. cube) and the dielectric has unit permeability, then resistance per km. in ohms

$$= 41{\cdot}2 \times 10^{-4} \sqrt{f}\left(\frac{1}{r_1} + \frac{1}{r_2}\right) \quad . \quad . \quad (7)$$

We can now write α in terms of feeder dimensions :

$$\alpha = \frac{41{\cdot}2 \times 10^{-4} \sqrt{f}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)}{2 \times 138 \log_{10} \frac{r_2}{r_1}}$$

$$= 14{\cdot}9 \times 10^{-6} \frac{\sqrt{f}\left(\frac{1}{r_1} + \frac{1}{r_2}\right)}{\log_{10} \frac{r_2}{r_1}} \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and hence, attenuation per km. in dbs.

$$= 1{\cdot}30 \times 10^{-4} \sqrt{f} \frac{\left(\frac{1}{r_1} + \frac{1}{r_2}\right)}{\log_{10} \frac{r_2}{r_1}} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

If we prefer to have an expression for the power efficiency of the line, that is, $\frac{\text{Power output}}{\text{Power input}}$, we can write (9) in the form

$$\log_{10} \eta = \frac{-\text{dbs.}}{10} = -1{\cdot}30 \times 10^{-5} \sqrt{f} \frac{\left(\frac{1}{r_1} + \frac{1}{r_2}\right)}{\log \frac{r_2}{r_1}} \quad . \quad . \quad (10)$$

this equation referring to a km. length of feeder.

Let us now assume the outer radius r_2 to be fixed, and investigate the effect of varying r_1 denoting the ratio by x. Then equation (9) may be written (for a fixed frequency).

$$\text{Attenuation in dbs. per km.} = y = \frac{k\,(x+1)}{\log_{10} x}$$

and this will be a minimum when $\dfrac{dy}{dx} = 0$

$$\text{i.e. when } \frac{k \log_{10} x - k\left(x+1\right)\dfrac{1}{x}}{\left(\log_{10} x\right)^2} = 0$$

$$\text{or } \log_{10} x = \frac{x+1}{x} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

This equation is solved by the value 3·6.

In the curve of Fig. 336 $\dfrac{x+1}{\log_{10} x}$ is plotted against x, so that the variation of attenuation with ratio may be seen.

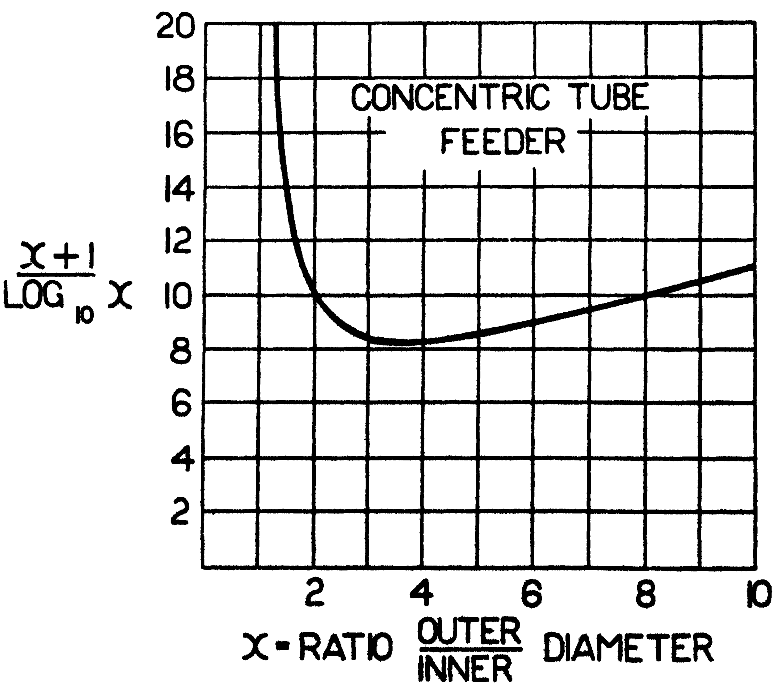


FIGURE 336.

The curve shows the great increase of attenuation if too small a ratio is used.

The resistance of parallel wire feeders may be worked out by considering each wire separately, using the formula (6) and putting $r_2 = \infty$. This method ignores earth current losses and the proximity effect of the other wire, this latter effect being usually small. By a similar procedure to that adopted above, the attenuation in dbs. per km.

$$= \frac{1{\cdot}30 \times 10^{-4}\sqrt{f}}{r \log_{10} \dfrac{d}{r}} \qquad . \quad . \quad . \quad (12)$$

and efficiency per km. is given by

$$\log_{10} \eta = \frac{-1{\cdot}30 \times 10^{-5}\sqrt{f}}{r \log_{10} \dfrac{d}{r}} \qquad . \quad . \quad (13)$$

When R and G are allowed for, the equations for voltage and current along an open-circuited line become :

$$E_x = E_G \frac{\cosh P(l - x)}{\cosh Pl} \qquad . \quad . \quad . \quad (14)$$

$$I_x = \frac{E_G}{Z_0} \cdot \frac{\sinh P(l - x)}{\cosh Pl} \qquad . \quad . \quad . \quad (15)$$

From (15), $$I_G = \frac{E_G}{Z_0} \tanh Pl \qquad . \quad . \quad . \quad . \quad . \quad (16)$$

and hence the impedance across the sending-end is given by

$$Z_{oc} = Z_0 \coth Pl \qquad . \quad . \quad . \quad . \quad . \quad (17)$$

Similarly, for a short-circuited line

$$E_x = E_G \frac{\sinh P(l - x)}{\sinh Pl} \qquad . \quad . \quad . \quad (18)$$

$$I_x = \frac{E_G}{Z_0} \cdot \frac{\cosh P(l - x)}{\sinh Pl} \qquad . \quad . \quad . \quad (19)$$

$$I_G = \frac{E_G}{Z_0} \coth Pl \qquad . \quad . \quad . \quad . \quad . \quad (20)$$

$$Z_{sc} = Z_o \tanh Pl \quad . \quad . \quad . \quad . \quad (21)$$

and, from (17) and (21), $Z_o = \sqrt{Z_{oc} Z_{sc}}$ . . . (22)

Equation (17) can be expanded into

$$Z_{oc} = Z_o \coth (\alpha + j\beta) l$$

$$= Z_o \frac{\cosh \alpha l \cos \beta l + j \sinh \alpha l \sin \beta l}{\sinh \alpha l \cos \beta l + j \cosh \alpha l \sin \beta l}$$

If $l = \frac{\lambda}{4}$, that is, $\beta l = \frac{\pi}{2}$, this becomes

$$Z_{oc} = Z_o \tanh \alpha l \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Similarly $Z_{sc} = Z_o \coth \alpha l$ . . . . . (24)

APPENDIX V

VALVES FOR SHORT AND ULTRA-SHORT WAVES

THE design of valves suitable for short-wave work calls for special attention (as has been made clear in the main text) and limitations in the standard design of valve developed for lower frequencies have necessitated changes in the mechanical structure of valves even when the transit time does not impose a limit to the frequency of operation.

A high mutual-conductance and input impedance are required and leads to anode, cathode and grid must have small inductance and capacity. In receiving valves it is desirable to bring the grid lead out separately as is done in the grid "top cap" type.

Suitable types of valves for short-wave and U.-S. wave receivers are shown in Table I.

The design of transmitting valves is even more difficult, as power considerations necessitate larger physical dimensions. Transmitting valves may be classified into four main groups, according to the way in which the heat produced at the anode is dissipated. Examples of each group are shown in Figs. 337 to 340, and in Table II is given a selection of typical valves suitable for short and ultra-short wave transmitters.

Group 1. Glass Valves with Radiation-Cooled Anodes. The air-cooled glass valve is generally speaking the cheapest valve to produce and it is suitable for most lower power installations. The anodes are usually of molydenum or carbon and in many cases the glass envelopes have all metallic deposits washed out after manufacture. At higher radio-frequencies, eddy currents circulate in these deposits and produce hot spots. If the washing-out process has not been done it is desirable to enclose the valve with a copper-strip or copper-gauze jacket. This has the effect of equalising the potential distribution over the glass surface, it not being necessary for such a jacket

Table I.

SOME TYPICAL RECEIVING VALVES.

VALVE			FILAMENT		CONDUCTANCE m A per Volt.		VOLTAGE		Interelectrode Cap. $\mu\mu F$			Max. Frequency Megacycles
Maker	Type	Class	Volts	Amperes	Mutual	Conversion	Anode	Screen	Grid / Anode	Grid / Cathode	Anode / Cathode	
MARCONI-OSRAM	KTW 61	Aligned Var-Mu Tetrode	6·3	·3	2·9	—	250	80	·0025	7·0	8·0	30
	KTW 63	Var-Mu Pentode	6·3	·3	1·5	—	250	125	—	—	—	—
	KTW 62	,, ,,	6·3	·3	2·9	—	250	100	—	—	—	—
	X 24	Triode-Hexode	2·0	·2	—	·250	150	60	·05	7·3	18·0	60
	KTZ 41	Triode-Tetrode	4·0	1·5	12·8 7·5	—	250	250	·008	14·0	10·5	60
	HA 1	(Acorn) Triode	6·3	·15	2·0	—	100	—	·6	1·4	·6	Detector to 700
	ZA 1	(Acorn) Pentode	4·0	·25	1·4	—	250	100	·007	3·0	3·0	300
	ZA 2	(Acorn) Pentode	6·3	·15	1·4	—	250	100	·007	3·0	3·0	300

to be connected to earth. The filament, anode and grid seals must be adequate to carry the high-frequency current through the valve in addition to the feed currents, and valves are often designed with the grid seal well removed from the filament seal. For valves up to some 200 watts dissipation it is,

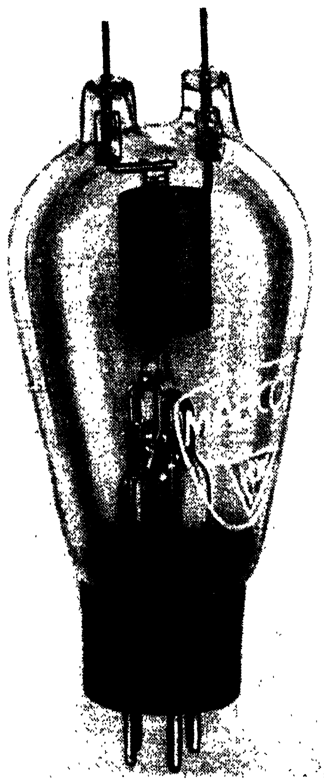
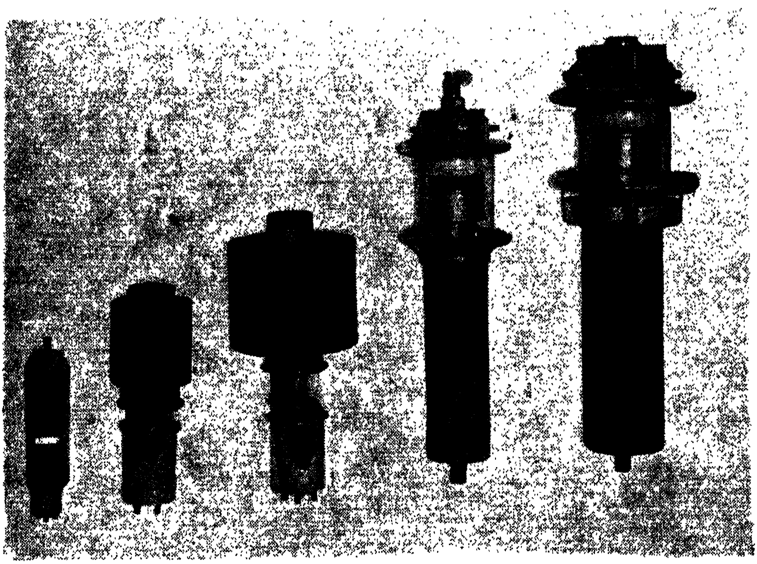


FIGURE 337.—Air-cooled D.E.T.12.

generally speaking, not necessary for forced air-draught to be used and it is not necessary to raise the filament voltage in steps when switching on.

Group 2. Anode Cooled by Air Convection. One of the disadvantages of the air-cooled glass valve is that the lead from the anode through the seal is, of necessity, of high inductance and small current-carrying capacity. With the

design of the air-cooled anode type of valve this objection has been overcome and heavy-current leads of low-inductance can be connected from the anode directly to the circuit. Using convection cooling this type of valve is commensurate with those of Group 1 as far as dissipation rating is concerned. It is often desirable to encourage convection and provide



1 2 3 4 5

FIGURE 338.—Examples of Pentode Transmitting Valves.
1. Air-cooled; 2 and 3. Air-cooled-anode (convection); 4 and 5. Water-cooled-anode.

increased radiation by bolting an attachment on to the anode proper, as shown in Fig. 338, 2 and 3.

Group 3. Anode Cooled by Forced Draught. The output from valves in Group 2 can be increased by the application of a forced draught around the anode and on to the other seals. This enables the dissipation-rating of the valve to be increased by some fifty per cent., Fig. 339 showing a typical example.

Group 4. Anode Cooled by Forced Water-Circulation. Water cooling is the most effective and hence all valves having

the highest dissipation-ratings will be of the water-cooled type. Valves capable of dissipating 150 kW are used successfully in short-wave transmitters. An interesting feature in the design of such a valve of the largest type is that the water-column thickness between anode and jacket is only about one-quarter

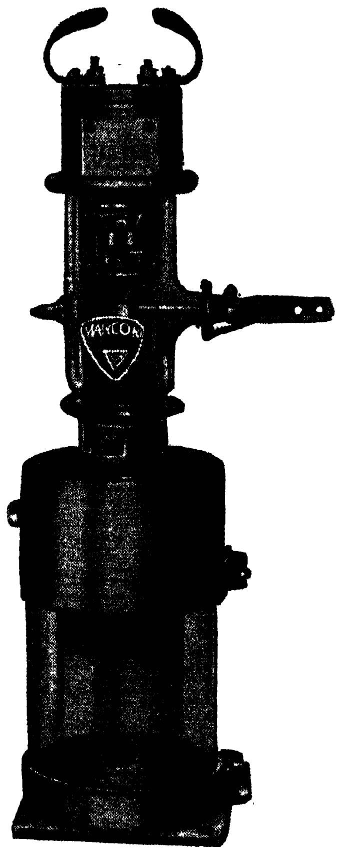


FIGURE 339.—Air-cooled-anode forced draught, A.C.T.10.

of an inch, the side of the valve being fluted or ribbed so as to encourage aeration of the water. Figs. 338 and 340 show examples of large and small water-cooled valves suitable for short waves.

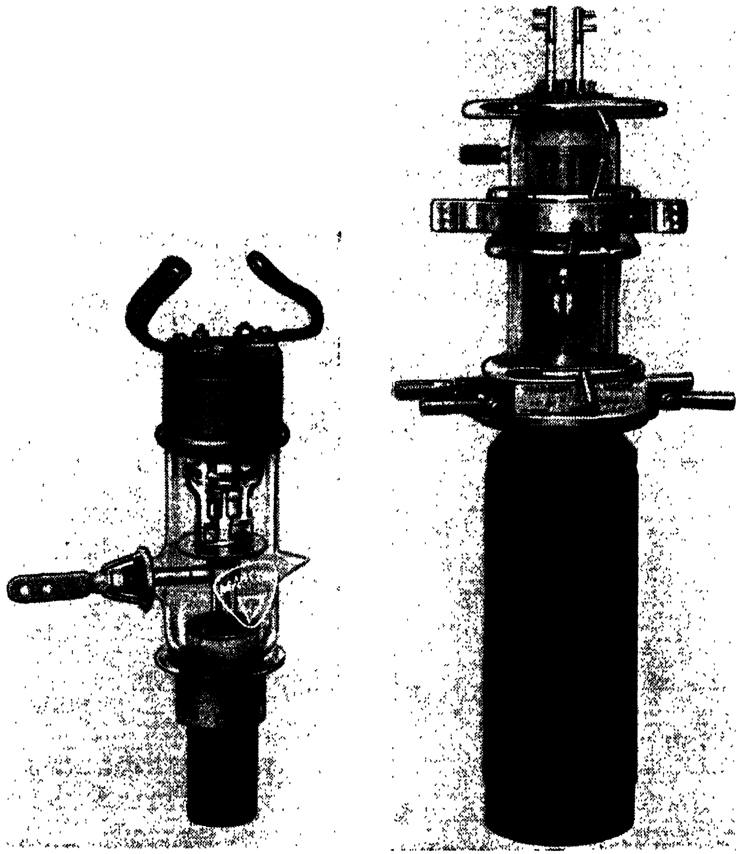
Valves for Very High Frequencies. In order to produce a valve suitable for very high-frequency amplification the

Table II.

SOME TYPICAL TRANSMITTING VALVES.

VALVE			FILAMENT			CONSTANTS			ANODE		Screen	Interelectrode Cap. $\mu\mu F$			Max. Frequency Mega-cycles
Maker	Type	Class	Volts	Amps	Emission Amps	μ	R_a	G_m	Volts	Dissipation Watts	Volts	Grid / Anode	Grid / Cathode	Anode / Cathode	
MARCONI–OSRAM	DET 20	Triode	6·3	·17	—	25	8,000	3·1	300	3·5	—	2·1	1·9	1·0	300
	DET 12	,,	7·5	3·15	1·5	10·3	5,400	1·9	1,250	50	—	3·0	2·1	·65	200
	DET 14	,,	7·5	3·0	—	21	9,500	2·2	1,500	55	—	3·75	4·0	1·5	120
	ACT 6	,,	10·0	1·6	·6	21	4,600	4·8	1,500 1,000 800	75	—	8·9	15·0	2·4	15 30 75
	CAT 10	,,	11·0	50	4·5	50	20,000	2·5	5,000	1,000	—	7·0	8·2	2·0	100
	CAT 15	,,	11·0	50	4·5	50	20,000	2·5	5,000	2,500	—	7·0	8·2	6·2	300
	CAT 17	,,	32·5	460	100	45	900	50	10,000	150 kW	—	51·0	10·1	6·6	22
	PT 6	Aligned Pentode	10·0	2·0	—	—	—	4·0	1,500	75	400	—	—	—	20
	PCA 21	,,	20	100	12	—	—	8·0	10,000	10 kW	1,500	—	—	—	60
	T.T. 9	Push-Pull Tetrode Door-Knob	12·6	·8	—	—	—	3·0	400	15 per Anode	—	·05	7·5	3·8	300
G.E.C. America	GL 880	Triode	12·6	320	—	—	—	—	7,500	20 kW	—	26	29	2·6	25
	GL 889	,,	11·0	125	—	—	—	—	8,500	5 kW	—	17·8	19·5	2·5	50
	GL 8002	,,	16·0	39	—	—	—	—	3,500	1,200	—	9·0	8·0	·5	150
R.C.A. America	834	Triode	—	—	—	10·5	—	—	1,200	50	—	9·0	2·2	·6	100
	826	Triode Door-Knob	—	—	—	31	—	—	—	50	—	—	—	—	250
	829	Push-Pull Tetrode Door-Knob	—	—	—	—	—	—	—	20 per Anode	—	—	—	—	200

R.C.A. developed the so-called "Acorn" type of valve shown in Fig. 342, introducing what is now known as the ring type of seal. This ring seal is made by pressing the edges of two cup-shaped glass envelopes together, the leading-in wires being



C.A.T.15. C.A.T.17.

FIGURE 340.—Water-cooled-anode.

disposed fan-wise around the edge of these cups. By such a design the various leads can be separated from one another as much as possible, thus reducing the capacity between them. The connections are also very direct, thus reducing the inductance, and the form of valve holder used adds little to the capacity

and inductance and is of a low-loss material. The principle on which the "Acorn" valve is designed is the so-called principle of similitude. It has been shown that if two valves are constructed which are similar, except that all the linear dimensions of the electrode structure of one valve are $\frac{1}{n}$th those of the other, then both valves will have similar values of anode resistance, amplification factor and mutual conductance. Inter-electrode capacitances and electron transit-time

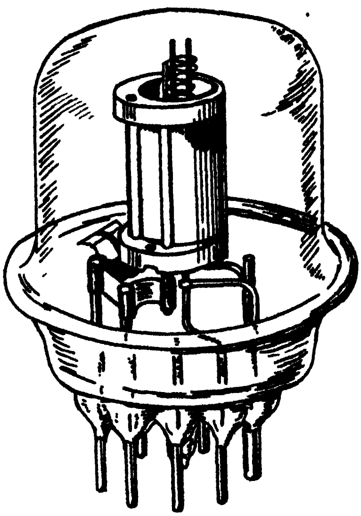


FIGURE 341.—"Door knob."

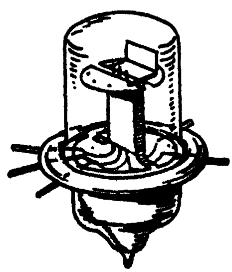


FIGURE 342.—"Acorn."

between the various electrodes will, however, be divided by this factor n. Of course, the anode dissipation and the cathode emission of the smaller valve will only be $\frac{1}{n^2}$ that of the large valve.

Valves employing this principle are built both for transmitting and receiving work and Figs. 341 and 342 indicate typical valves of each class. And selected examples of such types are shown in Table II.

Index

PRINTED BY JARROLD AND SONS, LIMITED, NORWICH